Wolfgang Demtröder

Laser Spectroscopy 1

Basic Principles

5th Edition

Springer

Wolfgang Demtröder

Laser Spectroscopy 1

Basic Principles

5th Edition

Springer

Prof. Dr. Wolfgang Demtröder
University of Kaiserslautern
Kaiserslautern, Germany

ISBN 978-3-642-53858-2 ISBN 978-3-642-53859-9 (eBook)
DOI 10.1007/978-3-642-53859-9

Library of Congress Control Number: 2014938249

Springer

Printed on acid-free paper

Springer is part of Springer Science+Business Media
www.springer.com

Chapter 4
Spectroscopic Instrumentation

This chapter is devoted to a discussion of instruments and techniques that are of fundamental importance for the measurements of wavelengths and line profiles, or for the sensitive detection of radiation. The optimum selection of proper equipment or the application of a new technique is often decisive for the success of an experimental investigation. Since the development of spectroscopic instrumentation has shown great progress in recent years, it is most important for any spectroscopist to be informed about the state-of-the-art regarding sensitivity, spectral resolving power, and signal-to-noise ratios attainable with modern equipment.

At first we discuss the basic properties of *spectrographs* and*monochromators*. Although for many experiments in laser spectroscopy these instruments can be replaced by monochromatic tunable lasers (Chap. 5 and Vol. 2, Chap. 1), they are still indispensible for the solution of quite a number of problems in spectroscopy.

Probably the most important instruments in laser spectroscopy are *interferometers*, which are applicable in various modifications to numerous problems. We therefore treat these devices in somewhat more detail. Recently, new techniques of measuring laser wavelengths with high accuracy have been developed; they are mainly based on interferometric devices. Because of their relevance in laser spectroscopy they will be discussed in a separate section.

Great progress has also been achieved in the field of low-level signal detection. Apart from new photomultipliers with an extended spectral sensivity range and large quantum efficiencies, new detection instruments have been developed such as image intensifiers, infrared detectors, charge-coupled devices (CCDs) or optical multichannel analyzers, which could move from classified military research into the open market. For many spectroscopic applications they prove to be extremely useful.

W. Demtröder, *Laser Spectroscopy 1*, DOI 10.1007/978-3-642-53859-9_4,

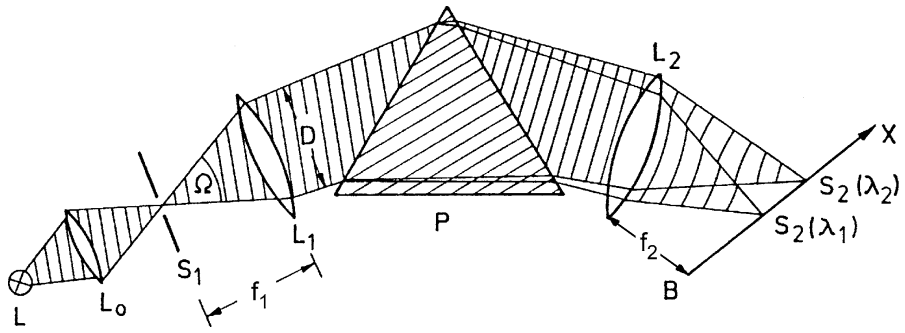

Figure 4.1 Prism spectrograph

4.1 Spectrographs and Monochromators

Spectrographs, the first instruments for measuring wavelengths, still hold their position in spectroscopic laboratories, particularly when equipped with modern accessories such as computerized microdensitometers or optical multichannel analyzers. Spectrographs are optical instruments that form images $S_2(\lambda)$ of the entrance slit S_1; the images are laterally separated for different wavelengths λ of the incident radiation (Fig. 2.17). This lateral dispersion is due to either spectral dispersion in prisms or diffraction on plane or concave reflection gratings.

Figure 4.1 depicts the schematic arrangement of optical components in a *prism spectrograph*. The light source L illuminates the entrance slit S_1, which is placed in the focal plane of the collimator lens L_1. Behind L_1 the parallel light beam passes through the prism P, where it is diffracted by an angle $\theta(\lambda)$ depending on the wavelength λ. The camera lens L_2 forms an image $S_2(\lambda)$ of the entrance slit S_1. The position $x(\lambda)$ of this image in the focal plane of L_2 is a function of the wavelength λ. The *linear dispersion* $\mathrm{d}x/\mathrm{d}\lambda$ of the spectrograph depends on the spectral dispersion $\mathrm{d}n/\mathrm{d}\lambda$ of the prism material and on the focal length of L_2.

When a reflecting diffraction grating is used to separate the spectral lines $S_2(\lambda)$, the two lenses L_1 and L_2 are commonly replaced by two spherical mirrors M_1 and M_2, which image the entrance slit either onto the exit slit S_2, or via the mirror M onto a CCD array in the plane of observation (Fig. 4.2). Both systems can use either photographic or photoelectric recording. According to the kind of detection, we distinguish between *spectrographs* and *monochromators*.

In spectrographs a charge-coupled device (CCD) diode array is placed in the focal plane of L_2 or M_2. The whole spectral range $\Delta\lambda = \lambda_1(x_1) - \lambda_2(x_2)$ covered by the lateral extension $\Delta x = x_1 - x_2$ of the diode array can be simultaneously recorded. The cooled CCD array can accumulate the incident radiant power over long periods (up to 20 h). CCD detection can be employed for both pulsed and cw light sources. The spectral range is limited by the spectral sensitivity of available CCD materials and covers the region between about 200–1000 nm.

Monochromators, on the other hand, use photoelectric recording of a selected small spectral interval. An exit slit S_2, selecting an interval Δx_2 in the focal plane B,

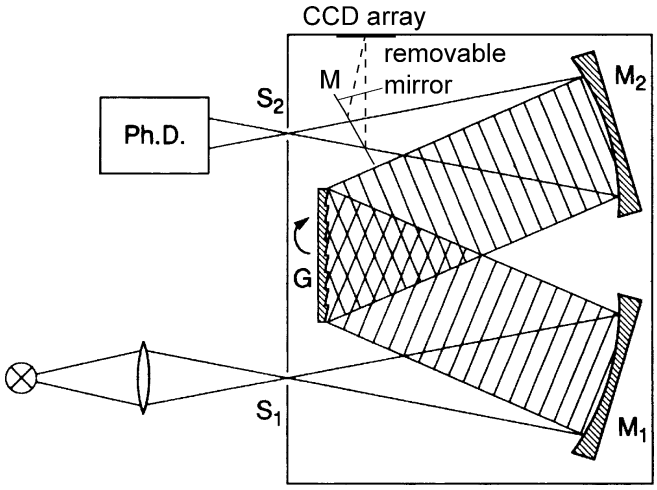

Figure 4.2 Grating monochromator

lets only the limited range $\Delta\lambda$ through to the photoelectric detector. Different spectral ranges can be detected by shifting S_2 in the x-direction. A more convenient solution (which is also easier to construct) turns the prism or grating by a gear-box drive, which allows the different spectral regions to be tuned across the fixed exit slit S_2. Modern devices uses a direct drive of the grating axis by step motors and measure the turning angle by electronic angle decoders. This avoids backlash of the driving gear. Unlike the spectrograph, different spectral regions are not detected simultaneously but successively. The signal received by the detector is proportional to the product of the area $h\Delta x_2$ of the exit slit with height h with the spectral intensity $\int I(\lambda)\mathrm{d}\lambda$, where the integration extends over the spectral range dispersed within the width Δx_2 of S_2.

Whereas the spectrograph allows the simultaneous measurement of a large region with moderate time resolution, photoelectric detection allows high time resolution but permits, for a given spectral resolution, only a small wavelength interval $\Delta\lambda$ to be measured at a time. With integration times below some minutes, photoelectric recording shows a higher sensitivity, while for very long detection times of several hours, photoplates may still be more convenient, although cooled CCD arrays currently allow integration times up to several hours.

In spectroscopic literature the name *spectrometer* is often used for both types of instruments. We now discuss the basic properties of spectrometers, relevant for laser spectroscopy. For a more detailed treatment see for instance [114–123].

4.1.1 Basic Properties

The selection of the optimum type of spectrometer for a particular experiment is guided by some basic characteristics of spectrometers and their relevance to the particular application. The basic properties that are important for all dispersive optical instruments may be listed as follows:

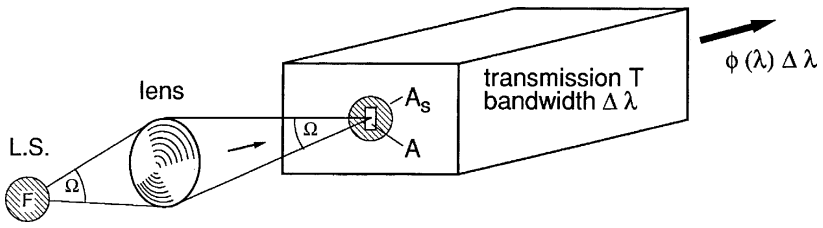

Figure 4.3 Light-gathering power of a spectrometer

a) Speed of a Spectrometer

When a light beam with spectral intensity $I^*(\lambda)$, cross section A_s and spectral radiation power

$$P_0(\lambda)\,\mathrm{d}\lambda = I^*_\lambda \cdot A_s \cdot \mathrm{d}\lambda \tag{4.1a}$$

within the solid angle $\mathrm{d}\Omega = 1\,\mathrm{sr}$ falls onto the entrance slit of a spectrometer with slit area $A < A_s$ and acceptance angel Ω, the power transmitted by the spectrometer is

$$P_t(\lambda)\,\mathrm{d}\lambda = P_0(A/A_s) \cdot T(\lambda) \cdot \Omega \cdot \mathrm{d}\lambda = I^*(\lambda) \cdot A \cdot T(\lambda) \cdot \Omega \cdot \mathrm{d}\lambda\ , \tag{4.1b}$$

where $T(\lambda)$ is the wavelength dependent transmission of the spectrometer.

The product $U = A\Omega$ is often named *étendue*. For the prism spectrograph the maximum solid angle of acceptance, $\Omega = F/f_1^2$, is limited by the effective area $F = hD$ of the parallel light beam transmitted through the prism, which represents the limiting aperture with height h and width D for the light beam (Fig. 4.1). For the grating spectrometer the sizes of the grating and mirrors limit the acceptance solid angle Ω.

Example 4.1
For a prism with height $h = 6\,\mathrm{cm}$, $D = 6\,\mathrm{cm}$, $f_1 = 30\,\mathrm{cm} \rightarrow D/f = 1:5$ and $\Omega = 0.04\,\mathrm{sr}$. With an entrance slit of $5 \times 0.1\,\mathrm{mm}^2$, the étendue is $U = 5 \times 10^{-3} \times 4 \times 10^{-2} = 2 \times 10^{-4}\,\mathrm{cm}^2\,\mathrm{sr}$.

In order to utilize the optimum speed, it is advantageous to image the light source onto the entrance slit in such a way that the acceptance angle Ω is fully used (Fig. 4.4). Although more radiant power from an extended source can pass the entrance slit by using a converging lens to reduce the source image on the entrance slit, the divergence is increased. The radiation outside the acceptance angle Ω cannot be detected, but may increase the background by scattering from lens holders and spectrometer walls.

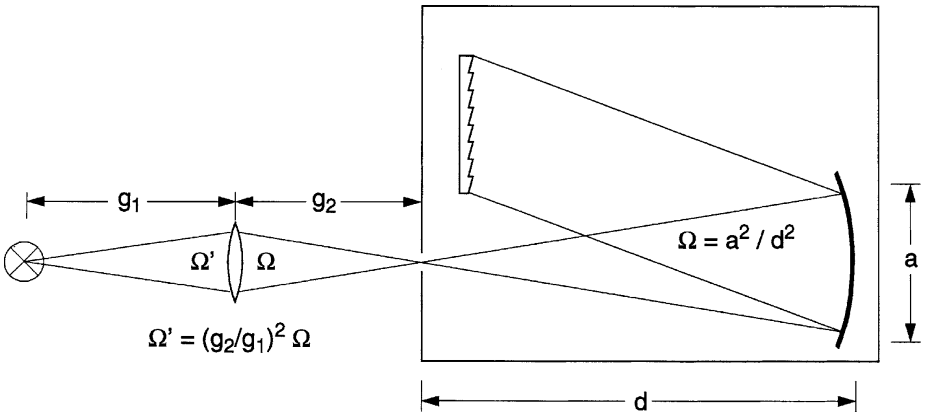

Figure 4.4 Optimized imaging of a light source onto the entrance slit of a spectrometer is achieved when the solid angle Ω' of the incoming light matches the acceptance angle $\Omega = (a/d)^2$ of the spectrometer

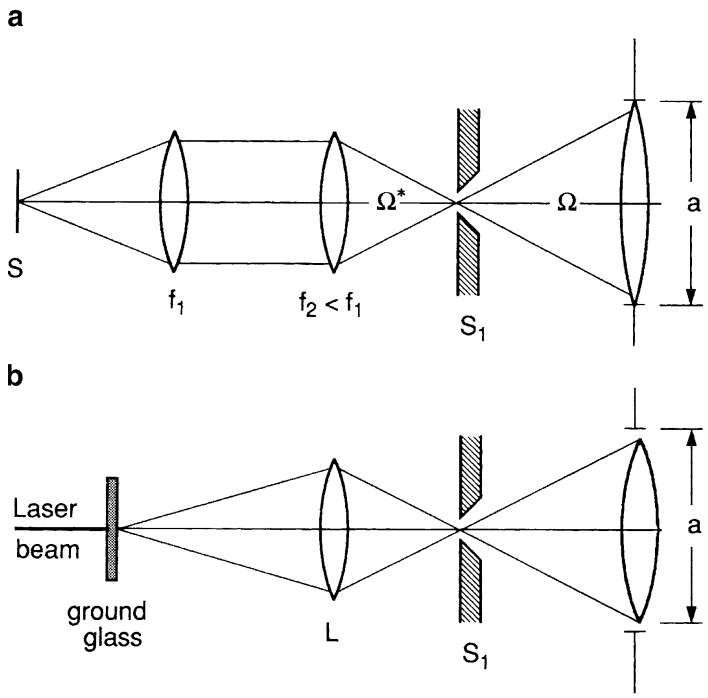

Figure 4.5 **a** Imaging of an extended light source onto the entrance slit of a spectrometer with $\Omega^* = \Omega$. **b** Correct imaging optics for laser wavelength measurements with a spectrometer. The laser light, scattered by the ground glass, forms the source that is imaged onto the entrance slit

Often the wavelength of lasers is measured with a spectrometer. In this case, it is not recommended to direct the laser beam directly onto the entrance slit, because the prism or grating would be not uniformely illuminated. This decreases the spectral resolution. Furthermore, the symmetry of the optical path with respect to the spectrometer axis is not guaranteed with such an arrangement, resulting in systematic errors of wavelengths measurements if the laser beam does not exactly coincide with the spectrometer axis. It is better to illuminate a ground-glass plate with the laser and to use the incoherently scattered laser light as a secondary source, which is imaged in the usual way (Fig. 4.5).

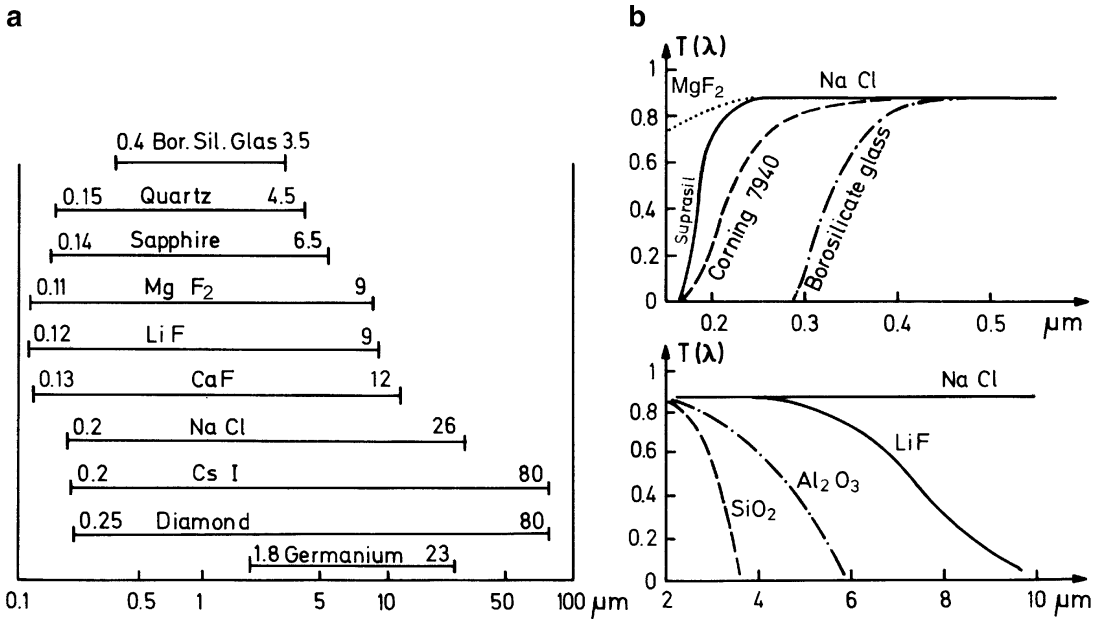

Figure 4.6 **a** Useful spectral ranges of different optical materials; and **b** transmittance of different materials with 1-cm thicknesses [118]b

b) Spectral Transmission

For prism spectrometers, the spectral transmission depends on the material of the prism and the lenses. Using fused quartz, the accessible spectral range spans from about 180 to 3000 nm. Below 180 nm (vacuum-ultraviolet region), the whole spectrograph must be evacuated, and lithium fluoride or calcium fluoride must be used for the prism and the lenses, although most VUV spectrometers are equipped with reflection gratings and mirrors.

In the infrared region, several materials (for example, CaF_2, NaCl, and KBr crystals) are transparent up to 30 µm, while CsI and diamond are transparent up to as high as 80 µm. (Fig. 4.6). However, because of the high reflectivity of metallic coated mirrors and gratings in the infrared region, grating spectrometers with mirrors are preferred over prism spectrographs.

Many vibrational–rotational transitions of molecules such as H_2O or CO_2 fall within the range 3–10 µm, causing selective absorption of the transmitted radiation. Infrared spectrometers therefore have to be either evacuated or filled with dry nitrogen. Because dispersion and absorption are closely related (see Sect. 2.6), prism materials with low absorption losses also show low dispersion, resulting in a limited spectral resolving power (see below).

Since the ruling or holographic production of high-quality gratings has reached a high technological standard, most spectrometers used today are equipped with diffraction gratings rather than prisms. The spectral transmission of grating spectrometers reaches from the VUV region into the far infrared. The design and the

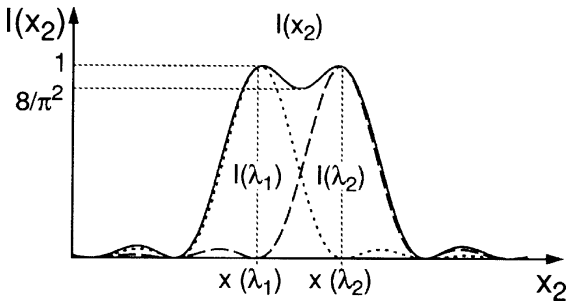

Figure 4.7 Rayleigh's criterion for the resolution of two nearly overlapping lines

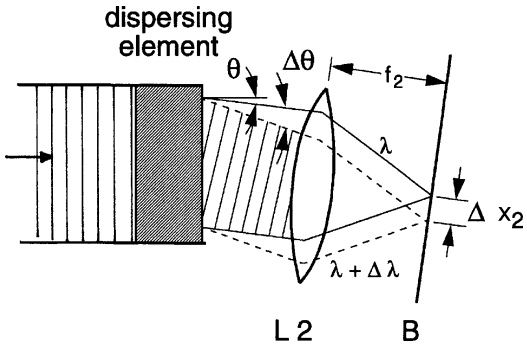

Figure 4.8 Angular dispersion of a parallel beam

coatings of the optical components as well as the geometry of the optical arrangement are optimized according to the specified wavelength region.

c) Spectral Resolving Power

The spectral resolving power of any dispersing instrument is defined by the expression

$$R = |\lambda/\Delta\lambda| = |\nu/\Delta\nu| \,, \tag{4.2}$$

where $\Delta\lambda = \lambda_1 - \lambda_2$ stands for the minimum separation of the central wavelengths λ_1 and λ_2 of two closely spaced lines that are considered to be just resolved. It is possible to recognize that an intensity distribution is composed of two lines with the intensity profiles $I_1(\lambda - \lambda_1)$ and $I_2(\lambda - \lambda_2)$ if the total intensity $I(\lambda) = I_1(\lambda - \lambda_1) + I_2(\lambda - \lambda_2)$ shows a pronounced dip between two maxima (Fig. 4.7). The intensity distribution $I(\lambda)$ depends, of course, on the ratio I_1/I_2 and on the profiles of both components. Therefore, the minimum resolvable interval $\Delta\lambda$ will differ for different profiles.

Lord Rayleigh introduced a criterion of resolution for diffraction-limited line profiles, where two lines are considered to be just resolved if the central diffraction maximum of the profile $I_1(\lambda - \lambda_1)$ coincides with the first minimum of $I_2(\lambda - \lambda_2)$ [116].

Let us consider the attainable spectral resolving power of a spectrometer. When passing the dispersing element (prism or grating), a parallel beam composed of two monochromatic waves with wavelengths λ and $\lambda + \Delta\lambda$ is split into two par-

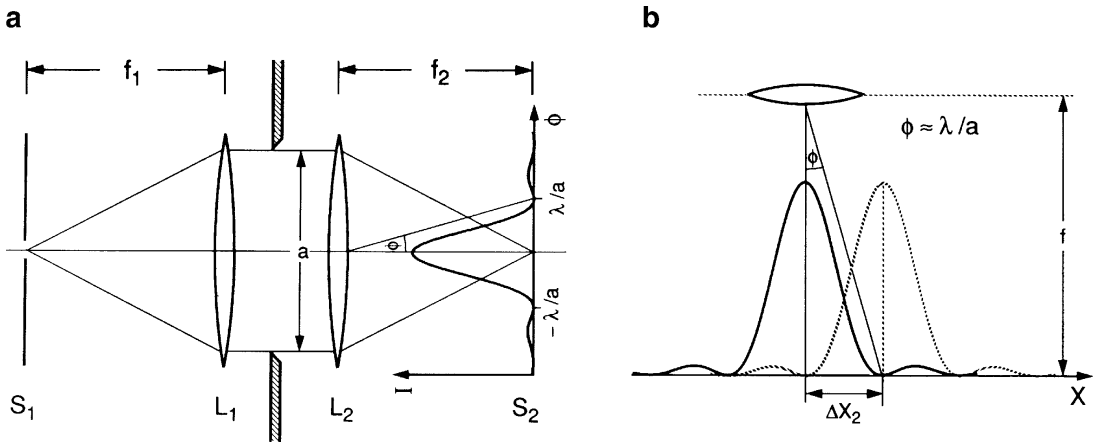

Figure 4.9 **a** Diffraction in a spectrometer by the limiting aperture with diameter a. **b** Limitation of spectral resolution by diffraction

tial beams with the angular deviations θ and $\theta + \Delta\theta$ from their initial direction (Fig. 4.8). The angular separation is

$$\Delta\theta = (\mathrm{d}\theta/\mathrm{d}\lambda)\Delta\lambda \ , \tag{4.3}$$

where $\mathrm{d}\theta/\mathrm{d}\lambda$ is called the *angular dispersion* [rad/nm]. Since the camera lens with focal length f_2 images the entrance slit S_1 into the plane B (Fig. 4.1), the distance Δx_2 between the two images $S_2(\lambda)$ and $S_2(\lambda + \Delta\lambda)$ is, according to Fig. 4.8,

$$\Delta x_2 = f_2 \Delta\theta = f_2 \frac{\mathrm{d}\theta}{\mathrm{d}\lambda}\Delta\lambda = \frac{\mathrm{d}x}{\mathrm{d}\lambda}\Delta\lambda \ . \tag{4.4}$$

The factor $\mathrm{d}x/\mathrm{d}\lambda$ is called the *linear dispersion* of the instrument. It is generally measured in mm/nm. In order to resolve two lines at λ and $\lambda + \Delta\lambda$, the separation Δx_2 in (4.4) has to be at least the sum $\delta x_2(\lambda) + \delta x_2(\lambda + \Delta\lambda)$ of the widths of the two slit images. Since the width δx_2 is related to the width δx_1 of the entrance slit according to geometrical optics by

$$\delta x_2 = (f_2/f_1)\delta x_1 \ , \tag{4.5}$$

the resolving power $\lambda/\Delta\lambda$ can be increased by decreasing δx_1. Unfortunately, there is a theoretical limitation set by diffraction. Because of the fundamental importance of this resolution limit, we discuss this point in more detail.

When a parallel light beam passes a limiting aperture with diameter a, a Fraunhofer diffraction pattern is produced in the plane of the focusing lens L_2 (Fig. 4.9). The intensity distribution $I(\phi)$ as a function of the angle ϕ with the optical axis of the system is given for $\phi \ll \pi/2 \Rightarrow \sin\phi \approx \phi$ by the well-known formula [116]

$$I(\phi) = I_0 \left(\frac{\sin(\pi \cdot (a/\lambda)\sin\phi)}{(a\pi\sin\phi)/\lambda} \right)^2 \simeq I_0 \left(\frac{\sin(a\pi\phi/\lambda)}{a\pi\phi/\lambda} \right)^2 \ . \tag{4.6}$$

The first two diffraction minima at $\phi = \pm\lambda/a \ll \pi$ are symmetrical to the central maximum (zeroth diffraction order) at $\phi = 0$. The intensity of the central diffraction maximum

$$I^{(0)} = \int_{-\lambda/a}^{+\lambda/a} I(\Phi)\,\mathrm{d}\Phi$$

contains about 90 % of the total intensity.

Even an infinitesimally small entrance slit therefore produces a slit image of width

$$\delta x_{\mathrm{s}}^{\mathrm{diffr}} = f_2(\lambda/a)\ , \tag{4.7}$$

defined as the distance between the central diffraction maximum and the first minimum, which is approximately equal to the FWHM of the central maximum.

According to the Rayleigh criterion, two equally intense spectral lines with wavelengths λ and $\lambda + \Delta\lambda$ are just resolved if the central diffraction maximum of $S_2(\lambda)$ coincides with the first minimum of $S_2(\lambda + \Delta\lambda)$ (see above). This means that their maxima are just separated by $\delta x_{\mathrm{S}}^{\mathrm{diffr}} = f_2(\lambda/a)$. From (4.6) one can compute that, in this case, both lines partly overlap with a dip of $(8/\pi^2)I_{\max} \approx 0.8 I_{\max}$ between the two maxima. The distance between the centers of the two slit images is then obtained from (4.7) (see Fig. 4.9b) as

$$\Delta x_2 = f_2(\lambda/a)\ . \tag{4.8a}$$

The separation of the two lines by dispersion (4.4) $\Delta x_2 = f_2(\mathrm{d}\theta/\mathrm{d}\lambda)\Delta\lambda$ has to be larger than this limit. This gives the fundamental limit on the resolving power

$$\boxed{|\lambda/\Delta\lambda| \le a(\mathrm{d}\theta/\mathrm{d}\lambda)\ ,} \tag{4.9}$$

which clearly depends only on the size a of the limiting aperture and on the angular dispersion of the instrument.

For a finite entrance slit with width b the separation Δx_2 between the central peaks of the two images $I(\lambda - \lambda_1)$ and $I(\lambda - \lambda_2)$ must be larger than (4.8a). We now obtain

$$\Delta x_2 \ge f_2\frac{\lambda}{a} + b\frac{f_2}{f_1}\ , \tag{4.8b}$$

in order to meet the Rayleigh criterion (Fig. 4.10). With $\Delta x_2 = f_2(\mathrm{d}\theta/\mathrm{d}\lambda)\Delta\lambda$, the smallest resolvable wavelength interval $\Delta\lambda$ is then

$$\Delta\lambda \ge \left(\frac{\lambda}{a} + \frac{b}{f_1}\right)\left(\frac{\mathrm{d}\theta}{\mathrm{d}\lambda}\right)^{-1}\ . \tag{4.10}$$

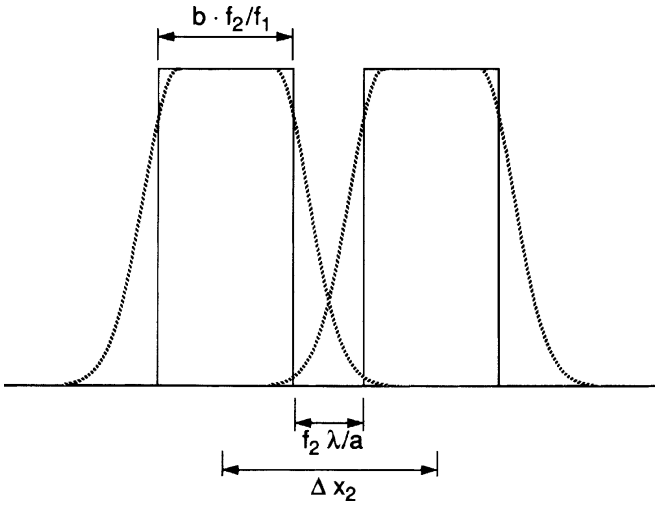

Figure 4.10 Intensity profiles of two monochromatic lines measured in the focal plane of L_2 with an entrance slit width $b \gg f_1 \cdot \lambda/a$ and a magnification factor f_2/f_1 of the spectrograph. *Solid line*: without diffraction; *dashed line*: with diffraction. The minimum resolvable distance between the line centers is $\Delta x_2 = f_2(b/f_1 + \lambda/a)$

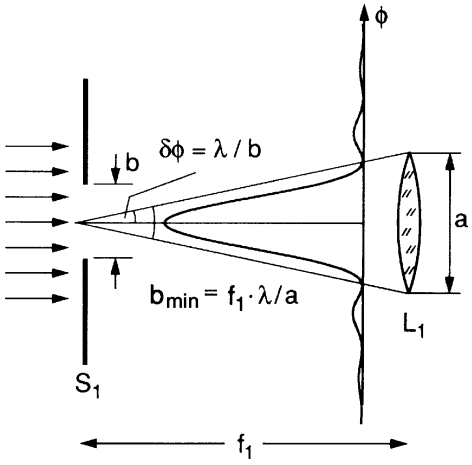

Figure 4.11 Diffraction by the entrance slit

Note The spectral resolution is limited, *not* by the diffraction due to the entrance slit, but by the diffraction caused by the much larger aperture a, determined by the size of the prism or grating.

Although it does not influence the spectral resolution, the much larger diffraction by the entrance slit imposes a limitation on the transmitted intensity at small slit widths. This can be seen as follows: when illuminated with parallel light, the entrance slit with width b produces a Fraunhofer diffraction pattern analogous to (4.6) with a replaced by b. The central diffraction maximum extends between the angles $\delta\phi = \pm\lambda/b$ (Fig. 4.11) and can completely pass the limiting aperture a only if $2\delta\phi$ is smaller than the acceptance angle a/f_1 of the spectrometer. This imposes a lower limit to the useful width $b_{\min}$ of the entrance slit,

$$b_{\min} \geq 2\lambda f_1/a \ . \tag{4.11}$$

In all practical cases, the incident light is divergent, which demands that the sum of the divergence angle and the diffraction angle has to be smaller than a/f and the minimum slit width b correspondingly larger.

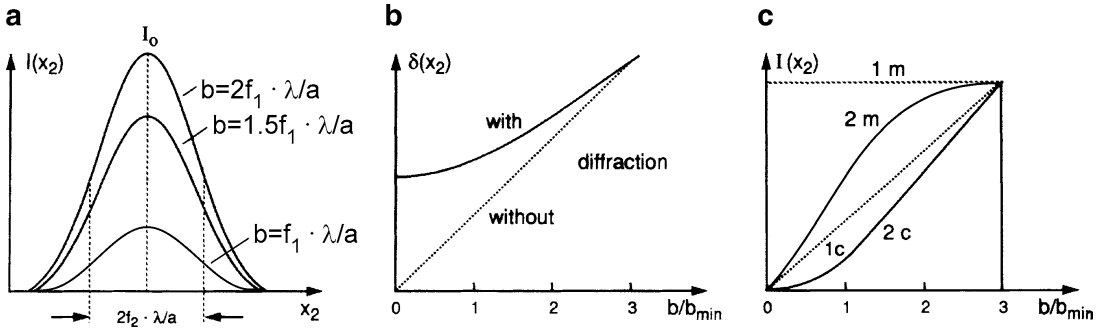

Figure 4.12 **a** Diffraction limited intensity distribution $I(x_2)$ in the plane B for different widths b of the entrance slit. **b** The width $\delta x_2(b)$ of the entrance slit image $S_2(x_2)$ with and without diffraction by the aperture a. **c** Intensity $I(x_2)$ in the observation plane as a function of entrance slit width b for a spectral continuum c and for a monochromatic spectral line (m) with diffraction (*solid curves* 2c and 2m) and without diffraction (*dashed curves* 1c and 1m)

Figure 4.12a illustrates the intensity distribution $I(x)$ in the plane B for different slit widths b. Figure 4.12b shows the dependence of the width $\Delta x_2(b)$ of the slit image S_2 on the entrance slit width b, taking into account the diffraction caused by the aperture a. This demonstrates that the resolution cannot be increased much by decreasing b below $b_{\min}$. The peak intensity $I(b)$ in the plane B is plotted in Fig. 4.12c as a function of the slit width b. According to (4.1b), the transmitted radiation flux $\phi(\lambda)$ depends on the product $U = A\Omega$ of the entrance slit area A and the acceptance angle $\Omega = (a/f_1)^2$. The flux in B would therefore depend linearly on the slit width b if diffraction were not present. This means that for monochromatic radiation the peak intensity [W/m^2] in the plane B should then be constant (*curve 1m*) although the transmitted power would increase linearly with b. For a spectral continuum it should decrease linearly with decreasing slit width (*curve 1c*). Because of the diffraction by S_1, the intensity decreases with the slit width b both for monochromatic radiation ($2m$) and for a spectral continuum ($2c$). Note the steep decrease for $b < b_{\min}$.

Substituting $b = b_{\min} = 2f\lambda/a$ into (4.10) yields the practical limit for $\Delta\lambda$ imposed by diffraction by S_1 and by the limiting aperture with width a

$$\Delta\lambda = 3f(\lambda/a)\mathrm{d}\lambda/\mathrm{d}x \; . \tag{4.12}$$

Instead of the theoretical limit (4.9) given by the diffraction through the aperture a, a smaller practically attainable resolving power is obtained from (4.12), which takes into account a finite minimum entrance slit width $b_{\min}$ imposed by intensity considerations and which yields:

$$\boxed{R = \lambda/\Delta\lambda = (a/3)\mathrm{d}\theta/\mathrm{d}\lambda \; .} \tag{4.13}$$

Example 4.2
For $a = 10\,\text{cm}$, $\lambda = 5 \times 10^{-5}\,\text{cm}$, $f = 100\,\text{cm}$, $\mathrm{d}\lambda/\mathrm{d}x = 1\,\text{nm/mm}$, with $b = 10\,\mu\text{m}, \rightarrow \Delta\lambda = 0.015\,\text{nm}$; with $b = 5\,\mu\text{m}, \rightarrow \Delta\lambda = 0.01\,\text{nm}$. However, from Fig. 4.12 one can see that the transmitted intensity with $b = 5\,\mu\text{m}$ is only 25 % of that with $b = 10\,\mu\text{m}$.

Note For photographic detection of line spectra, it is actually better to really use the lower limit $b_{\min}$ for the width of the entrance slit, because the density of the developed photographic layer depends only on the time-integrated spectral irradiance [W/m^2] rather than on the radiation power [W]. Increasing the slit width beyond the diffraction limit $b_{\min}$, in fact, does not significantly increase the density contrast on the plate, but does decrease the spectral resolution.

Using photoelectric recording, the detected signal depends on the radiation power $\phi_\lambda \mathrm{d}\lambda$ transmitted through the spectrometer and therefore increases with increasing slit width. In the case of completely resolved line spectra, this increase is proportional to the slit width b since $\phi_\lambda \propto b$. For continuous spectra it is even proportional to b^2 because the transmitted spectral interval $\mathrm{d}\lambda$ also increases proportional to b and therefore $\phi_\lambda \mathrm{d}\lambda \propto b^2$. Using diode arrays as detectors, the image $\Delta x_2 = (f_2/f_1)b$ should have the same width as one diode in order to obtain the optimum signal at maximum resolution.

The obvious idea of increasing the product of ΩA without loss of spectral resolution by keeping the width b constant but increasing the height h of the entrance slit is of limited value because imaging defects of the spectrometer cause a curvature of the slit image, which again decreases the resolution. Rays from the rim of the entrance slit pass the prism at a small inclination to the principal axis. This causes a larger angle of incidence α_2, which exceeds that of miniumum deviation. These rays are therefore refracted by a larger angle θ, and the image of a straight slit becomes curved toward shorter wavelengths (Fig. 4.13). Since the deviation in the plane B is equal to $f_2\theta$, the radius of curvature is of the same order of magnitude as the focal length of the camera lens and increases with increasing wavelength because of the decreasing spectral dispersion. In grating spectrometers, curved images of straight slits are caused by astigmatism of the spherical mirrors. The astigmatism can be partly compensated by using curved entrance slits [122]. Another solution is based on astigmatism-corrected imaging by using an asymmetric optical setup where the first mirror M_1 in Fig. 4.2 is placed at a distance $d_1 < f_1$ from the entrance slit and the exit slit at a distance $d_2 > f_2$ from M_2. In this arrangement [124] the grating is illuminated with slightly divergent light.

When the spectrometer is used as a monochromator with an entrance slit width b_1 and an exit slit width b_2, the power $P(t)$ recorded as a function of time while the grating is uniformly turned has a trapezoidal shape for $b_1 \gg b_{\min}$ (Fig. 4.14) with

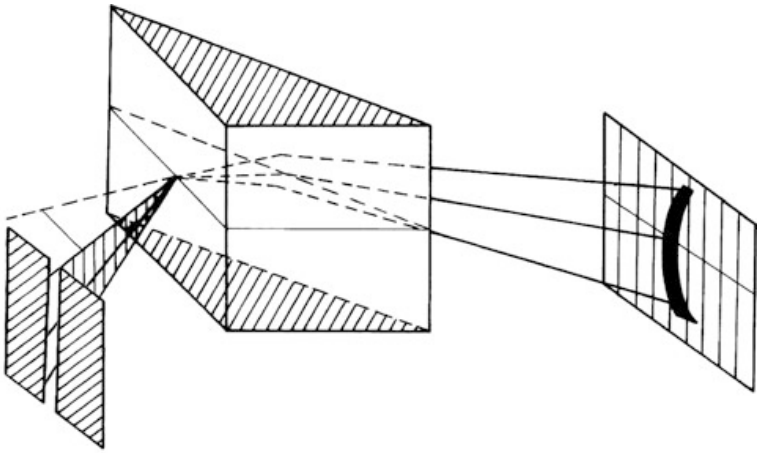

Figure 4.13 Curvature of the image of a straight entrance slit caused by astigmatic imaging errors

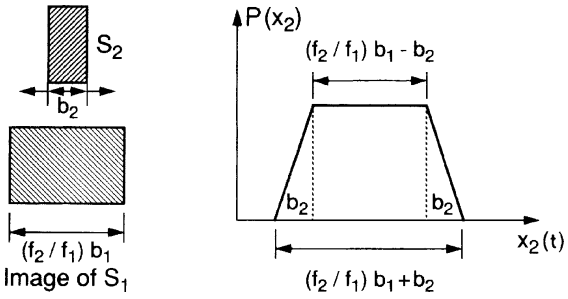

Figure 4.14 Signal profile $P(t) \propto P(x_2(t))$ at the exit slit of a monochromator with $b \gg b_{\min}$ and $b_2 < (f_2/f_1)b_1$ for monochromatic incident light with uniform turning of the grating

a baseline $(f_2/f_1)b_1 + b_2$. Optimum resolution at maximum transmitted power is achieved for $b_2 = (f_2/f_1)b_1$. The line profile $P(t) = P(x_2)$ then becomes a triangle.

d) Free Spectral Range

The free spectral range of a spectrometer is the wavelength interval $\delta\lambda$ of the incident radiation for which a one-valued relation exists between λ and the position $x(\lambda)$ of the entrance slit image. Two spectral lines with wavelengths λ_1 and $\lambda_2 = \lambda_1 \pm \delta\lambda$ cannot be distinguished without further information. This means that the wavelength λ measured by the instrument must be known beforehand with an uncertainty $\Delta\lambda < \delta\lambda$. While for prism spectrometers the free spectral range covers the whole region of normal dispersion of the prism material, for grating spectrometers $\delta\lambda$ is determined by the diffraction order m and decreases with increasing m (Sect. 4.1.3).

Interferometers, which are generally used in very high orders ($m = 10^4$–10^8), have a high spectral resolution but a small free spectral range $\delta\lambda$. For unambiguous wavelength determination they need a preselector, which allows one to measure the wavelength within the free spectral range $\delta\lambda$ of the high-resolution instrument (Sect. 4.2.4).

4.1.2 Prism Spectrometer

When passing through a prism, a light ray is refracted by an angle θ that depends on the prism angle ε, the angle of incidence α_1, and the refractive index n of the prism material (Fig. 4.15). We obtain from Fig. 4.15

$$\theta = \alpha_1 - \beta_1 + \alpha_2 - \beta_2 \ . \tag{4.14a}$$

Using the relation $\beta_1 + \beta_2 = \varepsilon$ for minimum refraction, where the light passes through the prism parallel to the baseline g of the prism we obtain

$$\theta = \alpha_1 + \alpha_2 - \varepsilon \tag{4.14b}$$

between the total deviation θ and the prism angle ε, we find the minimum refraction by differentiating:

$$\frac{\mathrm{d}\theta}{\mathrm{d}\alpha_1} = 1 + \frac{\mathrm{d}\alpha_2}{\mathrm{d}\alpha_1} = 0 \Rightarrow \mathrm{d}\alpha_1 = -\mathrm{d}\alpha_2 \ . \tag{4.14c}$$

From Snellius' law $\sin\alpha = n \ \sin\beta$ we obtain the derivatives:

$$\cos\alpha_1 \, \mathrm{d}\alpha_1 = n \ \cos\beta_1 \, \mathrm{d}\beta_1 \tag{4.14d}$$

$$\cos\alpha_2 \, \mathrm{d}\alpha_2 = n \ \cos\beta_2 \, \mathrm{d}\beta_2 \ . \tag{4.14e}$$

Because $\beta_1 + \beta_2 = \varepsilon \Rightarrow \mathrm{d}\beta_1 = -\mathrm{d}\beta_2$, the division of (4.14d) by (4.14e) yields

$$\frac{\cos\alpha_1 \, \mathrm{d}\alpha_1}{\cos\alpha_2 \, \mathrm{d}\alpha_2} = \frac{\cos\beta_1}{\cos\beta_2} \ .$$

For the minimum deviation θ with $\mathrm{d}\alpha_1 = -\mathrm{d}\alpha_2$ we get the result:

$$\frac{\cos\alpha_1}{\cos\alpha_2} = -\frac{\cos\beta_1}{\cos\beta_2} = -\left(\frac{1-\sin^2\beta_1}{1-\sin^2\beta_2}\right)^{1/2} \ . \tag{4.14f}$$

Squaring the equation yields

$$\frac{1-\sin^2\alpha_1}{1-\sin^2\alpha_2} = \frac{n^2-\sin^2\alpha_1}{n^2-\sin^2\alpha_2} \tag{4.14g}$$

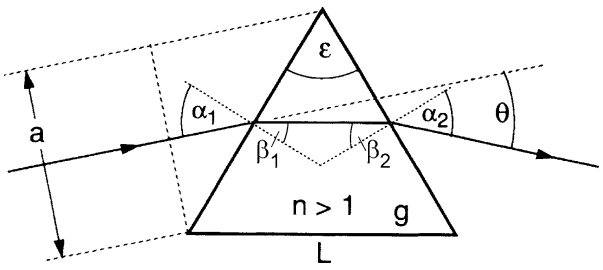

Figure 4.15 Refraction of light by a prism at minimum deviation where $\alpha_1 = \alpha_2 = \alpha$ and $\theta = 2\alpha - \varepsilon$

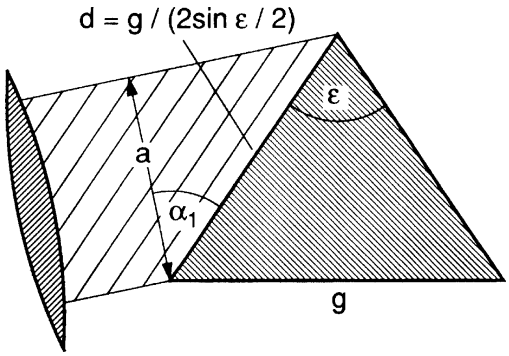

Figure 4.16 Limiting aperture in a prism spectrometer

which can only be fulfilled for $n \neq 1$ if $\alpha_1 = \alpha_2$. **The minimum deviation θ is obtained for symmetrical rays with** $\alpha_1 = \alpha_2 = \alpha$. The minimum deviation

$$\theta_{\min} = 2\alpha - \varepsilon \tag{4.14h}$$

is obtained when the ray passes the prism parallel to the base g. In this case, we derive from Snellius' law:

$$\sin\left(\frac{\theta_{\min} + \varepsilon}{2}\right) = \sin\alpha = n\ \sin\beta = n\ \sin(\varepsilon/2) \tag{4.14i}$$

$$\sin\left(\frac{\theta + \varepsilon}{2}\right) = n\ \sin(\varepsilon/2)\ . \tag{4.14j}$$

From (4.14j) the derivation $\mathrm{d}\theta/\mathrm{d}n = (\mathrm{d}n/\mathrm{d}\theta)^{-1}$ is

$$\frac{\mathrm{d}\theta}{\mathrm{d}n} = \frac{2\sin(\varepsilon/2)}{\cos[(\theta+\varepsilon)/2]} = \frac{2\sin(\varepsilon/2)}{\sqrt{1 - n^2\sin^2(\varepsilon/2)}}\ . \tag{4.15}$$

The *angular dispersion* $\mathrm{d}\theta/\mathrm{d}\lambda = (\mathrm{d}\theta/\mathrm{d}n)(\mathrm{d}n/\mathrm{d}\lambda)$ is therefore

$$\boxed{\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = \frac{2\sin(\varepsilon/2)}{\sqrt{1 - n^2\sin^2(\varepsilon/2)}}\frac{\mathrm{d}n}{\mathrm{d}\lambda}}\ . \tag{4.16}$$

This shows that the angular dispersion increases with the prism angle ε, *but does not depend on the size of the prism*.

For the deviation of laser beams with small beam diameters, small prisms can therefore be used without losing angular dispersion. In a prism spectrometer, however, the size of the prism determines the limiting aperture a and therefore the diffraction; it has to be large in order to achieve a large spectral resolving power (see previous section). For a given angular dispersion, an equilateral prism with $\varepsilon = 60°$ uses the smallest quantity of possibly expensive prism material. Because $\sin 30° = 1/2$, (4.16) then reduces to

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = \frac{\mathrm{d}n/\mathrm{d}\lambda}{\sqrt{1 - (n/2)^2}}\ . \tag{4.17}$$

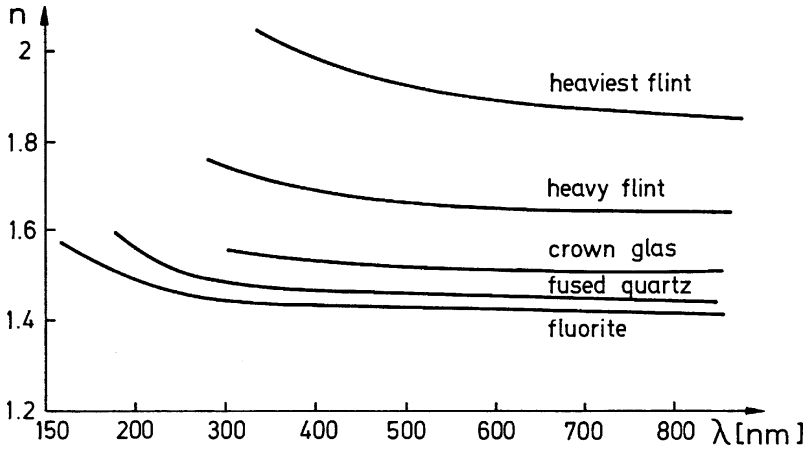

Figure 4.17 Refractive index $n(\lambda)$ for some prism materials

The diffraction limit for the resolving power $\lambda/\Delta\lambda$ according to (4.9) is

$$\lambda/\Delta\lambda \leq a(\mathrm{d}\theta/\mathrm{d}\lambda) \; .$$

The diameter a of the limiting aperture in a prism spectrometer is (Fig. 4.16)

$$a = \mathrm{d}\cos\alpha_1 = \frac{g\cos\alpha}{2\sin(\varepsilon/2)} \; . \tag{4.18}$$

Substituting $\mathrm{d}\theta/\mathrm{d}\lambda$ from (4.16) gives

$$\lambda/\Delta\lambda = \frac{g\cos\alpha_1}{\sqrt{1-n^2\sin^2(\varepsilon/2)}}\frac{\mathrm{d}n}{\mathrm{d}\lambda} \; . \tag{4.19}$$

At minimum deviation, (4.14a)–(4.14j) gives $n\sin(\varepsilon/2) = \sin(\theta+\varepsilon)/2 = \sin\alpha_1$ and therefore (4.19) reduces to

$$\lambda/\Delta\lambda = g(\mathrm{d}n/\mathrm{d}\lambda) \; . \tag{4.20a}$$

According to (4.20a), the theoretical maximum resolving power depends solely on the base length g and on the spectral dispersion of the prism material. Because of the finite slit width $b \geq b_{\min}$, the resolution reached in practice is somewhat lower. The corresponding resolving power can be derived from (4.11) to be at most

$$\boxed{R = \frac{\lambda}{\Delta\lambda} \leq \frac{1}{3}g\left(\frac{\mathrm{d}n}{\mathrm{d}\lambda}\right) \; .} \tag{4.20b}$$

The spectral dispersion $\mathrm{d}n/\mathrm{d}\lambda$ is a function of prism material and wavelength λ. Figure 4.17 shows dispersion curves $n(\lambda)$ for some materials commonly used for prisms. Since the refractive index increases rapidly in the vicinity of absorption

Table 4.1 Refractive index and dispersion of some materials used in prism spectrometers

Material	Useful spectral range [μm]	Refractive index n	Dispersion	$-\mathrm{d}n/\mathrm{d}\lambda$ [nm^{-1}]	
Glass (BK7)	0.35–3.5	1.516	4.6×10^{-5}	at	589 nm
		1.53	1.1×10^{-4}	at	400 nm
Heavy flint	0.4–2	1.755	1.4×10^{-4}	at	589 nm
		1.81	4.4×10^{-4}	at	400 nm
Fused quartz	0.15–4.5	1.458	3.4×10^{-5}	at	589 nm
		1.470	1.1×10^{-4}	at	400 nm
NaCl	0.2–26	1.79	6.3×10^{-3}	at	200 nm
		1.38	1.7×10^{-5}	at	20 μm
LiF	0.12–9	1.44	6.6×10^{-4}	at	200 nm
		1.09	8.6×10^{-5}	at	10 μm

lines, glass has a larger disperison in the visible and near-ultraviolet regions than quartz, which, on the other hand, can be used advantageously in the UV down to 180 nm. In the vacuum-ultraviolet range CaF, MgF, or LiF prisms are sufficiently transparent. Table 4.1 gives a summary of the optical characteristics and useful spectral ranges of some prism materials.

If achromatic lenses (which are expensive in the infrared and ultraviolet region) are not employed, the focal length of the two lenses decreases with the wavelength. This can be partly compensated by inclining the plane B against the principal axis in order to bring it at least approximately into the focal plane of L_2 for a large wavelength range (Fig. 4.1).

In Summary: The advantage of a prism spectrometer is the unambiguous assignment of wavelengths, since the position $S_2(\lambda)$ is a monotonic function of λ. Its drawback is the moderate spectral resolution. It is mostly used for survey scans of extended spectral regions.

Example 4.3

a) Suprasil (fused quartz) has a refractive index $n = 1.47$ at $\lambda = 400\,\mathrm{nm}$ and $\mathrm{d}n/\mathrm{d}\lambda = 1100\,\mathrm{cm}^{-1}$. This gives $\mathrm{d}\theta/\mathrm{d}\lambda = 1.6 \times 10^{-4}\,\mathrm{rad/nm}$. With a slitwidth $b_{\min} = 2f\lambda/a$ and $g = 5\,\mathrm{cm}$ we obtain from (4.20b) $\lambda/\Delta\lambda \leq 1830$. At $\lambda = 500\,\mathrm{nm} \Rightarrow \Delta\lambda \geq 0.27\,\mathrm{nm}$.

b) For heavy flint glass at 400 nm $n = 1.81$ and $\mathrm{d}n/\mathrm{d}\lambda = 4400\,\mathrm{cm}^{-1}$, giving $\mathrm{d}\theta/\mathrm{d}\lambda = 1.0 \times 10^{-3}\,\mathrm{rad/nm}$. This is about six times larger than that for quartz. With a focal length $f = 100\,\mathrm{cm}$ for the camera lens, one achieves a linear dispersion $\mathrm{d}x/\mathrm{d}\lambda = 1\,\mathrm{mm/nm}$ with a flint prism, but only 0.15 mm/nm with a quartz prism.

4.1.3 Grating Spectrometer

In a grating spectrometer (Fig. 4.2) the collimating lens L_1 is replaced by a spherical mirror M_1 with the entrance slit S_1 in the focal plane of M_1. The collimated parallel light is reflected by M_1 onto a reflection grating consisting of many straight grooves (about 10^5) parallel to the entrance slit. The grooves have been ruled onto an optically smooth glass substrate or have been produced by holographic techniques [125–131]. The whole grating surface is coated with a highly reflecting layer (metal or dielectric film). The light reflected from the grating is focused by the spherical mirror M_2 onto the exit slit S_2 or onto a photographic plate in the focal plane of M_2.

a) Basic Considerations

The many grooves, which are illuminated coherently, can be regarded as small radiation sources, each of them diffracting the light incident onto this small groove with a width $d \approx \lambda$ into a large range $\Delta r \approx \lambda/d$ of angles r around the direction of geometrical reflection (Fig. 4.18a). The total reflected light consists of a coherent superposition of these many partial contributions. Only in those directions where all partial waves emitted from the different grooves are in phase will constructive interference result in a large total intensity, while in all other directions the different contributions cancel by destructive interference.

Figure 4.18b depicts a parallel light beam incident onto two adjacent grooves. At an angle of incidence α to the grating normal (which is normal to the grating surface, but not necessarily to the grooves) one obtains constructive interference for those directions β of the reflected light for which the path difference $\Delta s = \Delta s_1 - \Delta s_2$ is an integer multiple m of the wavelength λ. With $\Delta s_1 = d \sin\alpha$ and $\Delta s_2 = d \sin\beta$

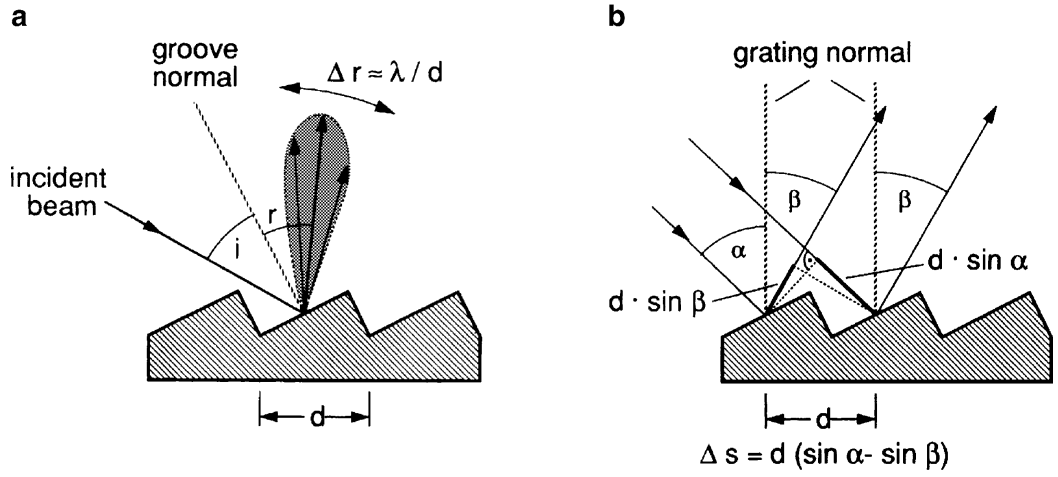

Figure 4.18 **a** Reflection of incident light from a single groove into the diffraction angle λ/d around the specular reflection angle $r = i$. **b** Illustration of the grating equation (4.21)

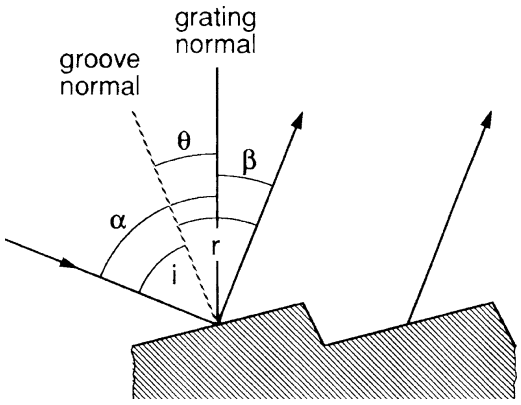

Figure 4.19 Illustration of the blaze angle θ

this yields the grating equation

$$\boxed{d(\sin\alpha \pm \sin\beta) = m\lambda\ ,} \tag{4.21}$$

the plus sign has to be taken if β and α are on the same side of the grating normal; otherwise the minus sign, which is the case shown in Fig. 4.18b.

The reflectivity $R(\beta, \theta)$ of a ruled grating depends on the diffraction angle β and on the blaze angle θ of the grating, which is the angle between the groove normal and the grating normal (Fig. 4.19). If the diffraction angle β coincides with the angle r of specular reflection from the groove surfaces, $R(\beta, \theta)$ reaches its optimum value R_0, which depends on the reflectivity of the groove coating. From Fig. 4.19 one infers for the case where α and β are on opposite sides of the grating normal, $i = \alpha - \theta$ and $r = \theta + \beta$, which yields, for specular reflection $i = r$, the condition for the optimum blaze angle θ

$$\theta = (\alpha - \beta)/2\ . \tag{4.22}$$

Because of the diffraction of each partial wave into a large angular range, the reflectivity $R(\beta)$ will not have a sharp maximum at $\beta = \alpha - 2\theta$, but will rather show a broad distribution around this optimum angle. The angle of incidence α is determined by the particular construction of the spectrometer, while the angle β for which constructive interference occurs depends on the wavelength λ. Therefore the blaze angle θ has to be specified for the desired spectral range and the spectrometer type.

In laser-spectroscopic applications the case $\alpha = \beta$ often occurs, which means that the light is reflected back into the direction of the incident light. For such an arrangement, called a *Littrow-grating mount* (shown in Fig. 4.20), the grating equation (4.21) for constructive interference reduces to

$$2d\,\sin\alpha = m\lambda\ . \tag{4.21a}$$

Maximum reflectivity of the Littrow grating is achieved for $i = r = 0 \rightarrow \theta = \alpha$ (Fig. 4.20b). The Littrow grating acts as a wavelength-selective reflector because light is only reflected if the incident wavelength satisfies the condition (4.21a).

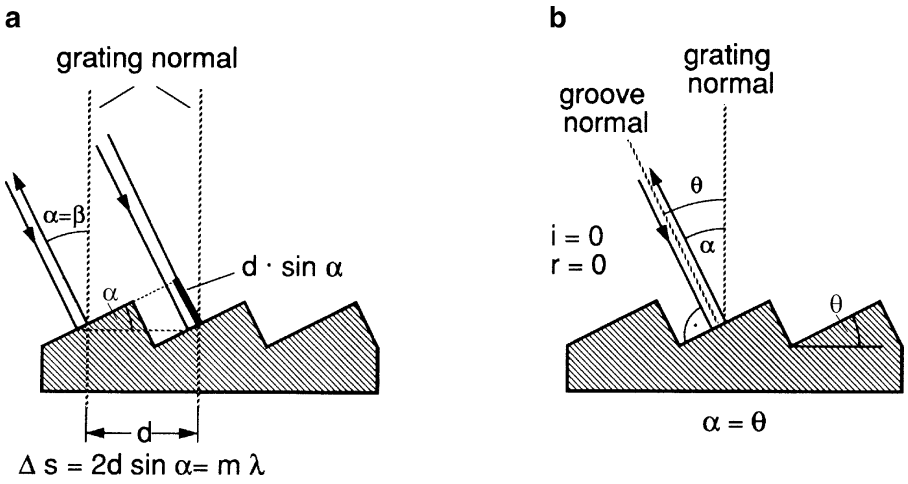

Figure 4.20 **a** Littrow mount of a grating with $\beta = \alpha$. **b** Illustration of blaze angle for a Littrow grating

For Littrow gratings used as wavelength-selective reflectors, it is desirable to have a high reflectivity in a selected order m and low reflections for all other orders. This can be achieved by selecting the width of the grooves and the blaze angle correctly. Because of diffraction by each groove with a width d, light can only reach angles β within the interval $\beta_0 \pm \lambda/d$ (Fig. 4.18a).

Example 4.4
With a blaze angle $\theta = \alpha = \beta = 30°$ and a step height $h = \lambda$, the grating can be used in second order, while the third order appears at $\beta = \beta_0 + 37°$. With $d = \lambda/\tan\theta = 2\lambda$, the central diffraction lobe extends only to $\beta_0 \pm 30°$, the intensity in the third order is very small.

b) Intensity Distribution of Reflected Light

We now examine the intensity distribution $I(\beta)$ of the reflected light when a monochromatic plane wave is incident onto an arbitrary grating.

According to (4.21) the path difference between partial waves reflected by adjacent grooves is $\Delta s = d(\sin\alpha \pm \sin\beta)$ and the corresponding phase difference is

$$\phi = \frac{2\pi}{\lambda}\Delta s = \frac{2\pi}{\lambda} d(\sin\alpha \pm \sin\beta) \ . \tag{4.23}$$

The superposition of the amplitudes reflected from all N grooves in the direction β gives the total reflected amplitude

$$A_{\mathrm{R}} = \sqrt{R}\sum_{m=0}^{N-1} A_{\mathrm{g}}\mathrm{e}^{\mathrm{i}m\phi} = \sqrt{R}A_{\mathrm{g}}\frac{1-\mathrm{e}^{\mathrm{i}N\phi}}{1-\mathrm{e}^{-\mathrm{i}\phi}} \ , \tag{4.24}$$

where $R(\beta)$ is the reflectivity of the grating, which depends on the reflection angle β, and A_{g} is the amplitude of the partial wave incident onto each groove. Because the intensity of the reflected wave is related to its amplitude by $I_{\mathrm{R}} = \epsilon_0 c A_{\mathrm{R}} A_{\mathrm{R}}^*$, see (2.30c), we find, with $\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x$, from (4.24),

$$I_{\mathrm{R}} = RI_0\frac{\sin^2(N\phi/2)}{\sin^2(\phi/2)} \quad \text{with } I_0 = c\epsilon_0 A_{\mathrm{g}} A_{\mathrm{g}}^* \ . \tag{4.25}$$

This intensity distribution is plotted in Fig. 4.21 for two different values of the total groove number N. Note that for real optical gratings $N \approx 10^5$! The principal maxima occur for $\phi = 2m\pi$, which is, according to (4.23), equivalent to the grating equation (4.21) and means that at a fixed angle α the path difference between partial beams from adjacent grooves is for certain angles β_m an integer multiple of the wavelength, where the integer m is called the *order of the interference*. The function (4.25) has $(N-1)$ minima with $I_{\mathrm{R}} = 0$ between two successive principal maxima. These minima occur at values of ϕ for which $N\phi/2 = \ell\pi$, $\ell = 1, 2, \ldots, N-1$, and mean that for each groove of the grating, another one can be found that emits light into the direction β with a phase shift π, such that all pairs of partial waves just cancel.

The line profile $I(\beta)$ of the principal maximum of order m around the diffraction angle β_m can be derived from (4.25) by substituting $\beta = \beta_m + \epsilon$. Because for large N, $I(\beta)$ is very sharply centered around β_m, we can assume $\epsilon \ll \beta_m$. With the relation

$$\sin(\beta_m + \epsilon) = \sin\beta_m\cos\epsilon + \cos\beta_m\sin\epsilon \sim \sin\beta_m + \epsilon\cos\beta_m \ ,$$

and because $(2\pi d/\lambda)(\sin\alpha + \sin\beta_m) = 2m\pi$, we obtain from (4.23)

$$\phi(\beta) = 2m\pi + 2\pi(d/\lambda)\epsilon\cos\beta_m = 2m\pi + \delta_1 \tag{4.26}$$

with

$$\delta_1 = 2\pi(d/\lambda)\epsilon\cos\beta_m \ll 1 \ .$$

Furthermore, (4.25) can be written as

$$I_{\mathrm{R}} = RI_0\frac{[\sin(Nm\pi + N\delta_1/2)]^2}{[\sin(m\pi + \delta_1/2)]^2} = RI_0\frac{\sin^2(N\delta_1/2)}{\sin^2(\delta_1/2)} \simeq RI_0N^2\frac{\sin^2(N\delta_1/2)}{(N\delta_1/2)^2} \ . \tag{4.27}$$

The first two minima with $I_R = 0$ on both sides of the central maximum at β_m are at

$$N\delta_1 = \pm 2\pi \Rightarrow \delta_1 = \pm 2\pi/N \ . \tag{4.28a}$$

From (4.26) we can now calculate the angular width $\Delta\beta$ of the central maximum around β_m: The first two minima on both sides of the intensity maximum appear for

$$\frac{2\pi d}{\lambda}\epsilon \cos\beta_m = \delta_1 = \pm\frac{2\pi}{N} \Rightarrow \tag{4.28b}$$

$$\epsilon_{1,2} = \frac{\pm\lambda}{Nd\cos\beta_m} \Rightarrow \Delta\beta = \frac{2\lambda}{N\,\mathrm{d}\cos\beta_m} \ . \tag{4.28c}$$

The intensity maximum of mth order therefore has a line profile (4.27) with a base full width $\Delta\beta = 2\lambda/(Nd\cos\beta_m)$. This corresponds to a diffraction pattern produced by an aperture with width $b = Nd\cos\beta_m$, which is just the size of the whole grating projected onto a plane, normal to the direction of β_m (Fig. 4.18).

Example 4.5
For $N \cdot d = 10\,\mathrm{cm}$, $\lambda = 5 \times 10^{-5}\,\mathrm{cm}$, $\beta_m = 45° \Rightarrow \cos\beta_m = \frac{1}{2}\sqrt{2} \Rightarrow$ $\varepsilon_{1/2} = 7 \times 10^{-6}\,\mathrm{rad} = 4 \times 10^{-4}\,°$.

Note According to (4.28a)–(4.28c) the full angular halfwidth $\Delta\beta = 2\epsilon$ of the interference maxima decreases as $1/N$, while according to (4.27) the peak intensity increases with the number of illuminated grooves proportional to $N^2 I_0$, where I_0 is the power incident onto a single groove. The area under the main maxima is therefore proportional to NI_0, which is due to the increasing concentration of light into the directions β_m. Of course, the incident power per groove decreases as $1/N$. The total reflected power is therefore independent of N.

The intensity of the $N - 2$ small side maxima, which are caused by incomplete destructive interference, decreases proportional to $1/N$ with increasing groove number N. Figure 4.21 illustrates this point for $N = 5$ and $N = 11$. For gratings used in practical spectroscopy with groove numbers of about 10^5, the reflected intensity $I_\mathrm{R}(\lambda)$ at a given wavelength λ has very sharply defined maxima only in those directions β_m, as defined by (4.21). The small side maxima are completely negligible at such large values of N, provided the distance d between the grooves is exactly constant over the whole grating area.

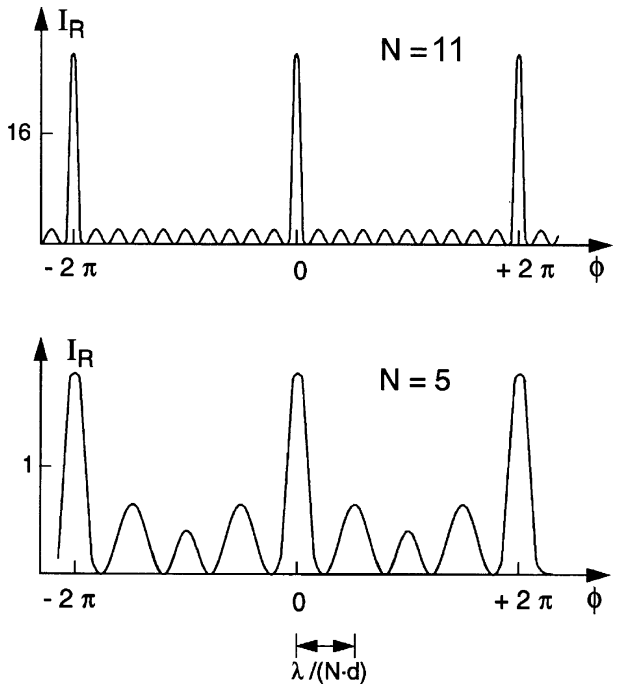

Figure 4.21 Intensity distribution $I(\beta)$ for two different numbers N of illuminated grooves. Note the different scales of the ordinates!

c) Spectral Resolving Power

Differentiating the grating equation (4.21) with respect to λ, we obtain at a given angle α the angular dispersion

$$\frac{\mathrm{d}\beta}{\mathrm{d}\lambda} = \frac{m}{\mathrm{d}\cos\beta} \,. \tag{4.29a}$$

Substituting $m/d = (\sin\alpha \pm \sin\beta)/\lambda$ from (4.21), we find

$$\frac{\mathrm{d}\beta}{\mathrm{d}\lambda} = \frac{\sin\alpha \pm \sin\beta}{\lambda\cos\beta} \,. \tag{4.29b}$$

This illustrates that the angular dispersion is determined solely by the angles α and β and *not by the number of grooves!* For the Littrow mount with $\alpha = \beta$ and the $+$ sign in (4.29b), we obtain

$$\frac{\mathrm{d}\beta}{\mathrm{d}\lambda} = \frac{2\tan\alpha}{\lambda} \,. \tag{4.29c}$$

The spectral resolving power can be immediately derived from (4.29a) and the base halfwidth $\Delta\beta = \epsilon = \lambda/(Nd\cos\beta)$ of the principal diffraction maximum (4.28a)–(4.28c), if we apply the Rayleigh criterion (see above) that two lines λ and $\lambda + \Delta\lambda$ are just resolved when the maximum of $I(\lambda)$ falls into the adjacent minimum for $I(\lambda + \Delta\lambda)$. This is equivalent to the condition

$$\frac{\mathrm{d}\beta}{\mathrm{d}\lambda}\Delta\lambda = \frac{\lambda}{Nd\cos\beta} \,,$$

or, inserting (4.29b):

$$\frac{\lambda}{\Delta\lambda} = \frac{Nd(\sin\alpha \pm \sin\beta)}{\lambda} , \tag{4.30}$$

which reduces with (4.21) to

$$\boxed{R = \frac{\lambda}{\Delta\lambda} = mN .} \tag{4.31}$$

The theoretical spectral resolving power equals the product of the diffraction order m with the total number N of illuminated grooves. If the finite slit width b_1 and the diffraction at limiting aperatures are taken into account, the practically achievable resolving power according to (4.13) is about 2–3 times lower.

Often it is advantageous to use the spectrometer in second order ($m = 2$), which increases the spectral resolution by a factor of 2 without losing much intensity, if the blaze angle θ is correctly choosen to satisfy (4.21) and (4.22) with $m = 2$.

Example 4.6

A grating with a ruled area of $10 \times 10\,\text{cm}^2$ and 10^3 grooves/mm allows in second order ($m = 2$) a theoretical spectral resolution of $R = 2 \times 10^5$. This means that at $\lambda = 500\,\text{nm}$ two lines that are separated by $\Delta\lambda = 2.5 \times 10^{-3}\,\text{nm}$ should be resolvable. Because of diffraction, the practical limit is $\Delta\lambda \approx 5 \times 10^{-3}\,\text{nm}$. The dispersion for $\alpha = \beta = 30°$ and a focal length $f = 1\,\text{m}$ is $\mathrm{d}x/\mathrm{d}\lambda = f\,\mathrm{d}\beta/\mathrm{d}\lambda = 2\,\text{mm/nm}$. With a slit width $b_1 = b_2 = 50\,\mu\text{m}$ a spectral resolution of $\Delta\lambda = 0.025\,\text{nm}$ can be achieved. In order to decrease the slit image width to $5 \times 10^{-3}\,\text{mm}$, the entrance slit width b has to be narrowed to $10\,\mu\text{m}$.

Lines around $\lambda = 1\,\mu\text{m}$ in the spectrum would appear in 1st order at the same angles β as lines with $\lambda = 500\,\text{mm}$ in 2nd order. They have to be suppressed by filters.

A special design is the so-called *echelle grating*, which has very widely spaced grooves forming right-angled steps (Fig. 4.22). The light is incident normal to the

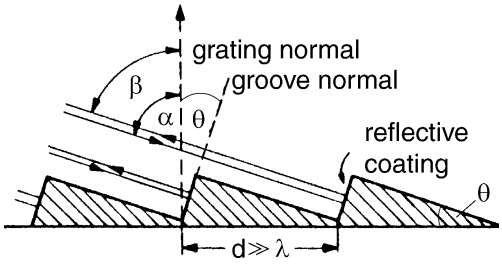

Figure 4.22 Echelle grating

small side of the grooves. The path difference between two reflected partial beams incident on two adjacent grooves with an angle of incidence $\alpha = 90° - \theta$ is $\Delta s = 2d \cos\theta$. The grating equation (4.21) gives for the angle β of the mth diffraction order

$$d(\cos\theta + \sin\beta) \approx 2d \cos\theta = m\lambda \ , \tag{4.32}$$

where β is close to $\alpha = 90° - \theta$.

With $d \gg \lambda$ the grating is used in a very high order ($m \simeq$ 10–100) and the resolving power is very high according to (4.31). Because of the larger distance d between the grooves, the achievable relative ruling accuracy is higher and large gratings (up to 30 cm) can be ruled. The disadvantage of the echelle is the small free spectral range $\delta\lambda = \lambda/m$ between successive diffraction orders.

Example 4.7
$N = 3 \times 10^4$, $d = 10\,\mu\text{m}$, $\theta = 30°$, $\lambda = 500\,\text{nm}$, $m = 34$. The spectral resolving power is $R = 10^6$, but the free spectral range is only $\delta\lambda = 15\,\text{nm}$. This means that the wavelengths λ and $\lambda + \delta\lambda$ overlap in the same direction β.

d) Grating Ghosts

Minute deviations of the distance d between adjacent grooves, caused by inaccuracies during the ruling process, may result in constructive interference from parts of the grating for "wrong" wavelengths. Such unwanted maxima, which occur for a given angle of incidence α into "wrong" directions β, are called *grating ghosts*. Although the intensity of these ghosts is generally very small, intense incident radiation at a wavelength λ_i may cause ghosts with intensities comparable to those of other weak lines in the spectrum. This problem is particularly serious in laser spectroscopy when the intense light at the laser wavelength, which is scattered by cell walls or windows, reaches the entrance slit of the monochromator.

In order to illustrate the problematic nature of achieving the ruling accuracy that is required to avoid these ghosts, let us assume that the carriage of the ruling engine expands by only 1 µm during the ruling of a $10 \times 10\,\text{cm}^2$ grating, e.g., due to temperature drifts. The groove distance d in the second half of the grating differs therefore from that of the first half by $5 \times 10^{-6} d$. With $N = 10^5$ grooves, the waves from the second half are then completely out of phase with those from the first half. The condition (4.21) is then fulfilled for different wavelengths in both parts of the grating, giving rise to unwanted wavelengths at the wrong positions β. Such ghosts are particularly troublesome in laser Raman spectroscopy (Vol. 2, Chap. 3) or low-level fluorescence spectroscopy, where very weak lines have to be detected in the presence of extremely strong excitation lines. The ghosts from these excitation lines

may overlap with the fluorescence or Raman lines and complicate the assignment of the spectrum.

e) Holographic Gratings

Although modern ruling techniques with interferometric length control have greatly improved the quality of ruled gratings [125–128] the most satisfactory way of producing completely ghost-free gratings is with holography. The production of holographic gratings proceeds as follows: a photosensitive layer on the grating's blank surface in the (x, y) plane is illuminated by two coherent plane waves with the wave vectors $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ ($|\boldsymbol{k}_1| = |\boldsymbol{k}_2|, \boldsymbol{k} = \{k_x, 0, k_z\}$), which form the angles α and β against the surface normal (Fig. 4.23). The intensity distribution of the superposition in the plane $z = 0$ of the photolayer consists of parallel dark and bright fringes imprinting an ideal grating into the layer, which becomes visible after developing the photoemulsion. The grating constant

$$d = \frac{\lambda/2}{\sin\alpha + \sin\beta}$$

depends on the wavelength $\lambda = 2\pi/|k|$ and on the angles α and β. Such holographic gratings are essentially free of ghosts. Their reflectivity R, however, is lower than that of ruled gratings and is furthermore strongly dependent on the polarization of the incident wave. This is due to the fact that holographically produced grooves are no longer planar, but have a sinusoidal surface and the "blaze angle" θ varies across each groove [130].

Summary: Summarizing the considerations above, we find that the grating acts as a wavelength-selective mirror, reflecting light of a given wavelength only into definite directions β_m, called the mth diffraction orders, which are defined by (4.21). The intensity profile of a diffraction order corresponds to the diffraction profile of a slit with width $b = Nd\cos\beta_m$ representing the size of the whole grating projection as seen in the direction β_m. *The spectral resolution* $\lambda/\Delta\lambda = mN =$

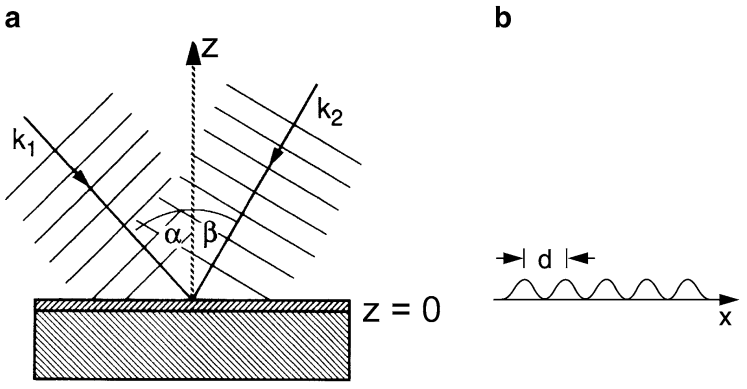

Figure 4.23 **a** Photographic production of a holographic grating; **b** surface of a holographic grating

$Nd(\sin\alpha+\sin\beta)/\lambda$ is therefore limited by the effective size of the grating measured in units of the wavelength.

For a more detailed discussion of special designs of grating monochromators, such as the concave gratings used in VUV spectroscopy, the reader is referred to the literature on this subject [125–131]. An excellent account of the production and design of ruled gratings can be found in [125].

4.2 Interferometers

For the investigation of the various line profiles discussed in Chap. 3, interferometers are preferentially used because, with respect to the spectral resolving power, they are superior even to large spectrometers. In laser spectroscopy the different types of interferometers not only serve to measure emission – or absorption – line profiles, but they are also essential devices for narrowing the spectral width of lasers, monitoring the laser linewidth, and controlling and stabilizing the wavelength of single-mode lasers (Chap. 5).

In this section we discuss some basic properties of interferometers with the aid of some illustrating examples. The characteristics of the different types of interferometers that are essential for spectroscopic applications are discussed in more detail. Since laser technology is inconceivable without dielectric coatings for mirrors, interferometers, and filters, an extra section deals with such dielectric multilayers. The extensive literature on interferometers [133–136] informs about special designs and applications.

4.2.1 Basic Concepts

The basic principle of all interferometers may be summarized as follows (Fig. 4.24). The indicent lightwave with intensity I_0 is divided into two or more partial beams with amplitudes A_k, which pass different optical path lengths $s_k = nx_k$ (where n is the refractive index) before they are again superimposed at the exit of the interferometer. Since all partial beams come from the same source, they are co-

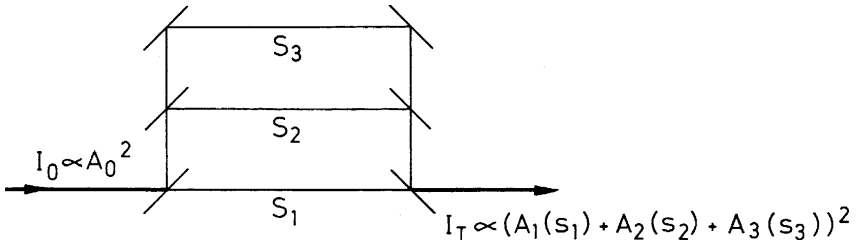

Figure 4.24 Schematic illustration of the basic principle for all interferometers

herent as long as the maximum path difference does not exceed the coherence length (Sect. 2.9). The total amplitude of the transmitted wave, which is the superposition of all partial waves, depends on the amplitudes A_k and on the phases $\phi_k = \phi_0 + 2\pi s_k/\lambda$ of the partial waves. *It is therefore sensitively dependent on the wavelength* λ.

The maximum transmitted intensity is obtained when all partial waves interfere constructively. This gives the condition for the optical path difference $\Delta s_{ik} = s_i - s_k$, namely

$$\Delta s_{ik} = m\lambda \ , \quad m = 1, 2, 3, \ldots \ . \tag{4.33}$$

The condition (4.33) for maximum transmission of the interferometer applies not only to a single wavelength λ but to all λ_m for which

$$\lambda_m = \Delta s/m \ , \quad m = 1, 2, 3, \ldots \ .$$

The wavelength interval

$$\delta\lambda = \lambda_m - \lambda_{m+1} = \frac{\Delta s}{m} - \frac{\Delta s}{m+1} = \frac{\Delta s}{m^2 + m} \tag{4.34a}$$

is called the *free spectral range* of the interferometer. With the mean wavelength $\bar{\lambda} = \frac{1}{2}(\lambda_m + \lambda_{m+1}) = \frac{1}{2}\Delta s(\frac{1}{m} + \frac{1}{m+1})$, we can write the free spectral range as:

$$\delta\lambda = \frac{2\bar{\lambda}}{2m+1} \ . \tag{4.34b}$$

It is more conveniently expressed in terms of frequency. With $\nu = c/\lambda$, (4.33) yields $\Delta s = mc/\nu_m$ and the free spectral frequency range

$$\boxed{\delta\nu = \nu_{m+1} - \nu_m = c/\Delta s \ ,} \tag{4.34c}$$

becomes independent of the order m.

Note It is important to realize that from one interferometric measurement alone one can only determine λ modulo $m \cdot \delta\lambda$ because all wavelengths $\lambda = \lambda_0 + m\delta\lambda$ are equivalent with respect to the transmission of the interferometer. One therefore has at first to measure λ within one free spectral range using other techniques before the absolute wavelength can be obtained with an interferometer.

Examples of devices in which only *two* partial beams interfere are the Michelson interferometer and the Mach–Zehnder interferometer. *Multiple*-beam interference is used, for instance, in the grating spectrometer, the Fabry–Perot interferometer, and in multilayer dielectric coatings of highly reflecting mirrors.

Some interferometers utilize the optical birefringence of specific crystals to produce two partial waves with mutually orthogonal polarization. The phase difference between the two waves is generated by the different refractive index for the two polarizations. An example of such a "polarization interferometer" is the *Lyot filter* [137] used in dye lasers to narrow the spectral linewidth (Sect. 4.2.11).

4.2.2 Michelson Interferometer

The basic principle of the Michelson interferometer (MI) is illustrated in Fig. 4.25. The incident plane wave

$$E = A_0 \mathrm{e}^{\mathrm{i}(\omega t - kx)}$$

is split by the beam splitter S (with reflectivity R and transmittance T) into two waves

$$E_1 = A_1 \exp\left[\mathrm{i}(\omega t - kx + \phi_1)\right] \quad \text{and} \quad E_2 = A_2 \exp\left[\mathrm{i}(\omega t - ky + \phi_2)\right] .$$

If the beam splitter has negligible absorption ($R + T = 1$), the amplitudes A_1 and A_2 are determined by $A_1 = \sqrt{T} A_0$ and $A_2 = \sqrt{R} A_0$ with $A_0^2 = A_1^2 + A_2^2$.

After being reflected at the plane mirrors M_1 and M_2, the two waves are superimposed in the plane of observation B. In order to compensate for the dispersion that beam 1 suffers by passing twice through the glass plate of beam splitter S, often an appropriate compensation plate P is placed in one side arm of the interferometer. The amplitudes of the two waves in the plane B are $\sqrt{TR} A_0$, because each wave has been transmitted and reflected once at the beam splitter surface S. The phase difference ϕ between the two waves is

$$\phi = \frac{2\pi}{\lambda} 2(\mathrm{SM}_1 - \mathrm{SM}_2) + \Delta\phi \, , \tag{4.35}$$

where $\Delta\phi$ accounts for additional phase shifts that may be caused by reflection. The total complex field amplitude in the plane B is then

$$E = \sqrt{RT} A_0 \mathrm{e}^{\mathrm{i}(\omega t + \phi_0)} (1 + \mathrm{e}^{\mathrm{i}\phi}) \, . \tag{4.36}$$

The detector in B cannot follow the rapid oscillations with frequency ω but measures the time-averaged intensity $\bar{I}$, which is, according to (2.30c),

$$\begin{aligned} \bar{I} &= \tfrac{1}{2} c\epsilon_0 A_0^2 RT (1 + \mathrm{e}^{\mathrm{i}\phi})(1 + \mathrm{e}^{-\mathrm{i}\phi}) = c\epsilon_0 A_0^2 RT (1 + \cos\phi) \\ &= \tfrac{1}{2} I_0 (1 + \cos\phi) \quad \text{for } R = T = \tfrac{1}{2} \text{ and } I_0 = \tfrac{1}{2} c\epsilon_0 A_0^2 \, . \end{aligned} \tag{4.37}$$

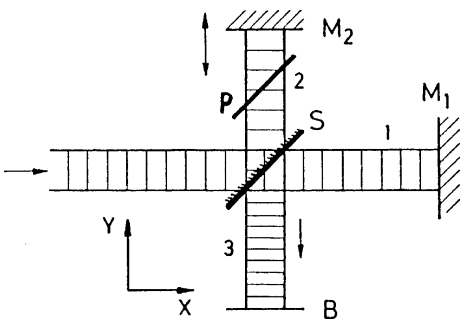

Figure 4.25 Two-beam interference in a Michelson interferometer

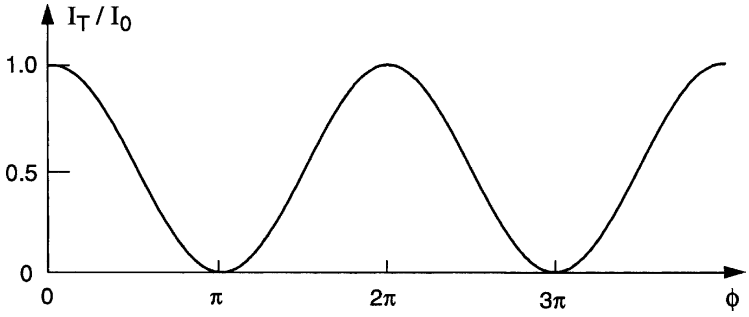

Figure 4.26 Intensity transmitted through the Michelson interferometer as a function of the phase difference ϕ between the two interfering beams for $R = T = 0.5$

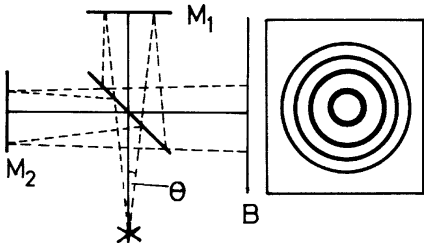

Figure 4.27 Circular fringe pattern produced by the MI with divergent incident light

If mirror M_2 (which is mounted on a carriage) moves along a distance Δy, the optical path difference changes by $\Delta s = 2n\Delta y$ (n is the refractive index between S and M_2) and the phase difference ϕ changes by $2\pi\Delta s/\lambda$. Figure 4.26 shows the intensity $I_T(\phi)$ in the plane B as a function of ϕ for a monochromatic incident plane wave. For the maxima at $\phi = 2m\pi$ ($m = 0, 1, 2, \ldots$), the transmitted intensity I_T becomes equal to the incident intensity I_0, which means that the transmission of the interferometer is $T_I = 1$ for $\phi = 2m\pi$. In the minima for $\phi = (2m + 1)\pi$ the transmitted intensity I_T is zero! The incident plane wave is being reflected back into the source.

This illustrates that the MI can be regarded either as a wavelength-dependent filter for the transmitted light, or as a wavelength-selective reflector. In the latter function it is often used for mode selection in lasers (Fox–Smith selector, Sect. 5.4.3).

For divergent incident light the path difference between the two waves depends on the inclination angle (Fig. 4.27). In the plane B an interference pattern of circular fringes, concentric to the symmetry axis of the system, is produced. Moving the mirror M_2 causes the ring diameters to change. The intensity behind a small aperture in the plane B, centered around the interferometer axis still follows approximately the function $I(\phi)$ in Fig. 4.26. With parallel incident light but slightly tilted mirrors M_1 or M_2, the interference pattern consists of parallel fringes, which move into a direction perpendicular to the fringes when Δs is changed.

The MI can be used for absolute wavelength measurements by counting the number N of maxima in B when the mirror M_2 is moved along a known distance Δy. The wavelength λ is then obtained from

$$\lambda = 2n\Delta y/N \ .$$

This technique has been applied to very precise determinations of laser wavelengths (Sect. 4.4).

The MI may be described in another equivalent way, which is quite instructive. Assume that the mirror M_2 in Fig. 4.25 moves with a constant velocity $v = \Delta y/\Delta t$. A wave with frequency ω and wave vector $\boldsymbol{k}$ incident perpendicularly on the moving mirror suffers a Doppler shift

$$\Delta\omega = \omega - \omega' = 2\boldsymbol{k}\cdot\boldsymbol{v} = (4\pi/\lambda)v \ , \tag{4.38}$$

on reflection.

Inserting the path difference $\Delta s = \Delta s_0 + 2vt$ and the corresponding phase difference $\phi = (2\pi/\lambda)\Delta s$ into (4.37) gives, with (4.38) and $\Delta s_0 = 0$,

$$\bar{I} = \tfrac{1}{2}\bar{I}_0(1 + \cos\Delta\omega t) \quad \text{with } \Delta\omega = 2\omega v/c \ . \tag{4.39}$$

We recognize (4.39) as the time-averaged beat signal, obtained from the superposition of two waves with frequencies ω and $\omega' = \omega - \Delta\omega$, giving the averaged intensity of

$$\bar{I} = I_0(1 + \cos\Delta\omega t)\overline{\cos^2[(\omega' + \omega)t/2]x} = \tfrac{1}{2}\bar{I}_0(1 + \cos\Delta\omega t) \ .$$

Note that the frequency $\omega = (c/v)\Delta\omega/2$ of the incoming wave can be measured from the beat frequency $\Delta\omega$, provided the velocity v of the moving mirror is known. The MI with uniformly moving mirror M_2 can be therefore regarded as a device that transforms the high frequency ω (10^{14}–$10^{15}\,\mathrm{s}^{-1}$) of an optical wave into an easily accessible rf-range $(v/c)\omega$.

Example 4.8
$v = 3\,\mathrm{cm/s} \rightarrow (v/c) = 10^{-10}$. The frequency $\omega = 3 \times 10^{15}\,\mathrm{Hz}$ ($\lambda = 0.6\,\mu\mathrm{m}$) is transformed to $\Delta\omega = 6 \times 10^5\,\mathrm{Hz} \simeq \Delta\nu \sim 100\,\mathrm{kHz}$.

The maximum path difference Δs that still gives interference fringes in the plane B is limited by the coherence length of the incident radiation (Sect. 2.9). Using spectral lamps, the coherence length is limited by the Doppler width of the spectral lines and is typically a few centimeters. With stabilized single-mode lasers, however, coherence lengths of several kilometers can be achieved. In this case, the maximum path difference in the MI is, in general, not restricted by the source but by technical limits imposed by laboratory facilities.

The attainable path difference Δs can be considerably increased by an *optical delay line*, placed in one arm of the interferometer (Fig. 4.28). It consists of a pair of mirrors, M_3, M_4, which reflect the light back and forth many times. In order to keep diffraction losses small, spherical mirrors, which compensate by collimation the divergence of the beam caused by diffraction, are preferable. With a stable mounting

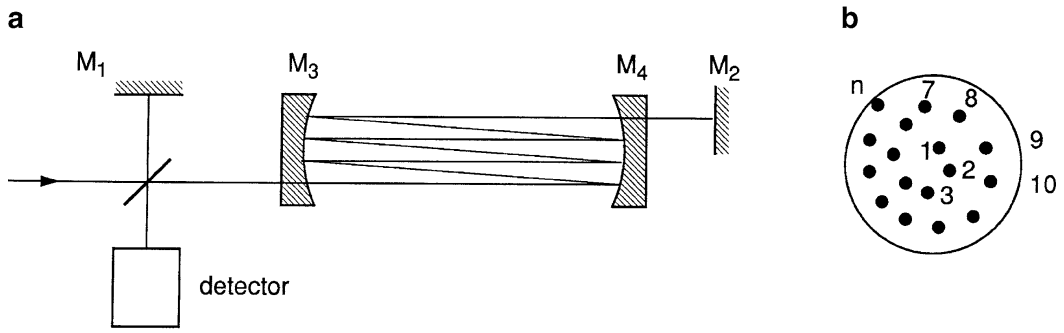

Figure 4.28 Michelson interferometer with optical delay line allowing a large path difference between the two interfering beams: **a** schematic arrangement; **b** spot positions of the reflected beams on mirror M_3

of the whole interferometer, optical path differences up to 350 m could be realized [138], allowing a spectral resolution of $\nu/\Delta\nu \simeq 10^{11}$. This was demonstrated by measuring the linewidth of a HeNe laser oscillating at $\nu = 5\times10^{14}$ Hz as a function of discharge current. The accuracy obtained was better than 5 kHz.

For gravitational-wave detection [139], a MI with side arms of about 1-km length has been built where the optical path difference can be increased to $\Delta s > 100$ km by using highly reflective spherical mirrors and an ultrastable solid-state laser with a coherence length of $\Delta s_c \gg \Delta s$ (see Vol. 2, Sect. 9.8, Sect. 9.8.3) [140].

4.2.3 *Fourier Spectroscopy*

When the incoming wave consists of two components with frequencies ω_1 and ω_2, the interference pattern varies with time according to

$$\begin{aligned}\bar{I}(t) &= \tfrac{1}{2}\bar{I}_{10}[1+\cos 2\omega_1(v/c)t] + \tfrac{1}{2}I_{20}[1+\cos 2\omega_2(v/c)t] \\ &= \bar{I}_0\{1+\cos[(\omega_1-\omega_2)vt/c]\cos[(\omega_1+\omega_2)vt/c]\}\,, \end{aligned} \tag{4.40}$$

where we have assumed $I_{10} = I_{20} = I_0$. This is a beat signal, where the amplitude of the interference signal at $(\omega_1+\omega_2)(v/c)$ is modulated at the difference frequency $(\omega_1-\omega_2)v/c$ (Fig. 4.29). From the sum

$$(\omega_1+\omega_2)+(\omega_1-\omega_2) = 2\omega_1$$

we obtain the frequency ω_1, and from the difference

$$(\omega_1+\omega_2)-(\omega_1-\omega_2) = 2\omega_2$$

the frequency ω_2.

The spectral resolution can roughly be estimated as follows: if Δy is the path difference traveled by the moving mirror in Fig. 4.25, the number of interference

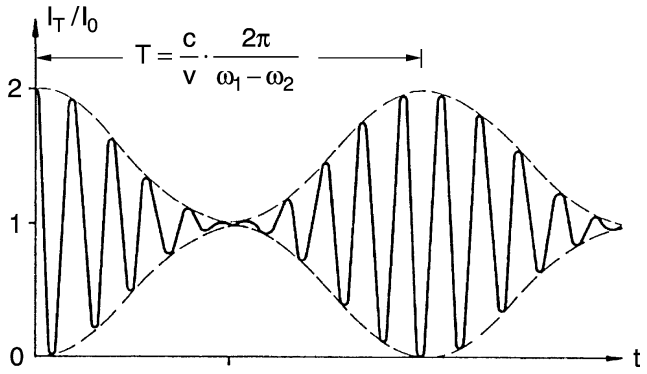

Figure 4.29 Interference signal behind the MI with uniformly moving mirror M_2 when the incident wave consists of two components with frequencies ω_1 and ω_2 and equal amplitudes

maxima that are counted by the detector is $N_1 = 2\Delta y/\lambda_1$ for an incident wave with the wavelength λ_1, and $N_2 = 2\Delta y/\lambda_2$ for $\lambda_2 < \lambda_1$. The two wavelengths can be clearly distinguished when $N_2 \geq N_1 + 1$. This yields with $\lambda_1 = \lambda_2 + \Delta\lambda$ and $\Delta\lambda \ll \lambda$ for the spectral resolving power

$$\frac{\lambda}{\Delta\lambda} = \frac{2\Delta y}{\lambda} = N = \frac{\Delta s}{\lambda} \quad \text{with } \lambda = (\lambda_1 + \lambda_2)/2 \text{ and } N = \tfrac{1}{2}(N_1 + N_2)\ . \tag{4.41a}$$

The equivalent consideration in the frequency domain follows. In order to determine the two frequencies ω_1 and ω_2, one has to measure at least over one modulation period

$$T = \frac{c}{v}\frac{2\pi}{\omega_1 - \omega_2} = \frac{c}{v}\frac{1}{\nu_1 - \nu_2}\ .$$

The frequency difference that can be resolved is then

$$\Delta\nu = \frac{c}{vT} = \frac{c}{\Delta s} = \frac{c}{N\lambda} \Rightarrow \frac{\Delta\nu}{c/\lambda} = \frac{1}{N} \text{ or } \frac{\nu}{\Delta\nu} = N = \frac{\Delta s}{\lambda}\ . \tag{4.41b}$$

The spectral resolving power $\lambda/\Delta\lambda$ of the Michelson interferometer equals the maximum path difference $\Delta s/\lambda$ measured in units of the wavelength λ.

Example 4.9

a) $\Delta y = 5\,\text{cm}$, $\lambda = 10\,\mu\text{m} \rightarrow N = 10^4 \Rightarrow \nu/\Delta\nu = 10^4$,

b) $\Delta y = 100\,\text{cm}$, $\lambda = 0.5\,\mu\text{m} \rightarrow N = 4 \times 10^6 \Rightarrow \Delta\nu = 2.5 \times 10^{-7}\nu$ where the latter example can be realized only with lasers that have a sufficiently large coherence length (Sect. 4.4).

c) $\lambda_1 = 10\,\mu\text{m}$, $\lambda_2 = 9.8\,\mu\text{m} \rightarrow (\nu_2 - \nu_1) = 6 \times 10^{11}\,\text{Hz}$; with $v = 1\,\text{cm/s} \rightarrow T = 50\,\text{ms}$. The minimum measuring time for the resolution of the two spectral lines is 50 ms, and the minimum path difference $\Delta s = vT = 5 \times 10^{-2}\,\text{cm} = 500\,\mu\text{m}$.

When the incoming radiation is composed of several components with frequencies ω_k, the total amplitude in the plane B of the detector is the sum of all interference amplitudes (4.36),

$$\boldsymbol{E} = \sum_k A_k \mathrm{e}^{\mathrm{i}(\omega_k t + \phi_{0k})}(1 + \mathrm{e}^{\mathrm{i}\phi_k}) \; . \tag{4.42a}$$

A detector with a large time constant compared with the maximum period $1/(\omega_i - \omega_k)$ does not follow the rapid oscillations of the amplitude at frequencies ω_k or at the difference frequencies $(\omega_i - \omega_k)$, but gives a signal proportional to the sum of the intensities I_k in (4.37). We therefore obtain for the time-dependent total intensity

$$\bar{I}(t) = \sum_k \tfrac{1}{2}\bar{I}_{k0}(1 + \cos\phi_k) = \sum_k \tfrac{1}{2}\bar{I}_{k0}(1 + \cos\Delta\omega_k t) \; , \tag{4.42b}$$

where the audio frequencies $\Delta\omega_k = 2\omega_k v/c$ are determined by all differences between the frequencies ω_k of the components and by the velocity v of the moving mirror. Measurements of these frequencies $\Delta\omega_k$ allows one to reconstruct the spectral components of the incoming wave with frequencies ω_k (Fourier transform spectroscopy [141, 142]).

Since the path-difference $\Delta s(t) = v \cdot t$ is a continuous function of the time t, the sum in (4.42b) can be replaced by the integral

$$I(t) = \int I(\omega)\, \cos(\omega\, \Delta s/c)\, \mathrm{d}\omega \; . \tag{4.42c}$$

The Fourier-transform of the measured intensity $I(t)$ in (4.42c) gives the wanted spectrum

$$I(\omega) = \int I(t) \cdot \cos(\omega \cdot vt/c)\, \mathrm{d}t \; . \tag{4.42d}$$

The main advantage of Fourier spectroscopy is the fact, that all spectral intervals $\mathrm{d}\omega$ with the intensity $I(\omega)\,\mathrm{d}\omega$ are measured simultaneously in contrast to classical spectroscopy with a monochromator where the different spectral intervals are measured subsequently. If a spectrum consisting of N spectral intervals $\Delta\omega$ (where $\Delta\omega$ is the spectral interval which can be resolved by the wavelength-selecting instrument) is measured in a time T with a monochromator, Fourier spectroscopy can obtain this spectrum in the shorter time $T/\sqrt{N}$ with the same signal-to-noise ratio.

4.2.4 Mach–Zehnder Interferometer

Analogous to the Michelson interferometer, the Mach–Zehnder interferometer is based on the two-beam interference by amplitude splitting of the incoming wave.

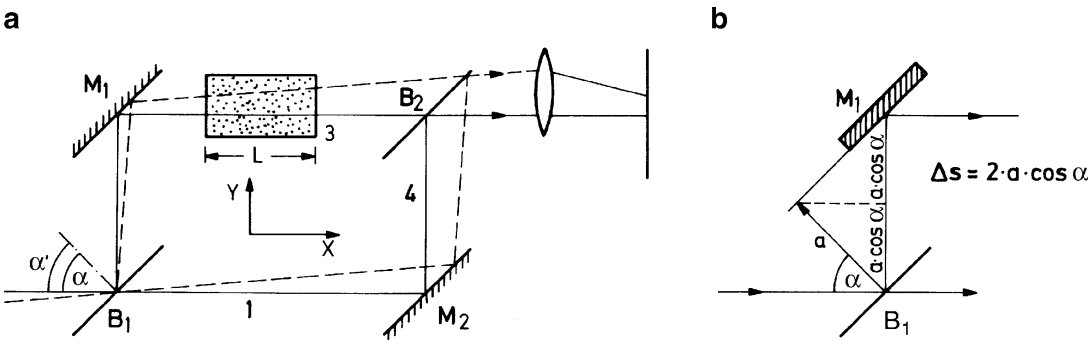

Figure 4.30 Mach–Zehnder interferometer: **a** schematic arrangement, **b** path difference between the two parallel beams

The two waves travel along different paths with a path difference $\Delta s = 2a \cos\alpha$ (Fig. 4.30b). Inserting a transparent object into one arm of the interferometer alters the optical path difference between the two beams. This results in a change of the interference pattern, which allows a very accurate determination of the refractive index of the sample and its local variation. The Mach–Zehnder interferometer may be regarded therefore as a sensitive refractometer.

If the beam splitters B_1, B_2 and the mirrors M_1, M_2 are all strictly parallel, the path difference between the two split beams does not depend on the angle of incidence α because the path difference between the beams 1 and 3 is exactly compensated by the same path length of beam 4 between M_2 and B_2 (Fig. 4.30a). This means that the interfering waves in the symmetric interferometer (without sample) experience the same path difference on the solid path as on the dashed path in Fig. 4.30a. Without the sample, the total path difference is therefore zero; it is $\Delta s = (n-1)L$ *with* the sample having the refractive index n in one arm of the interferometer.

Expanding the beam on path 3 gives an extended interference-fringe pattern, which reflects the local variation of the refractive index. Using a laser as a light source with a large coherence length, the path lengths in the two interferometer arms can be made different without losing the contrast of the interference pattern (Fig. 4.31). With a beam expander (lenses L_1 and L_2), the laser beam can be expanded up to 10–20 cm and large objects can be tested. The interference pattern can either be photographed or may be viewed directly with the naked eye or with a television camera [143]. Such a laser interferometer has the advantage that the laser beam diameter can be kept small everywhere in the interferometer, except between the two expanding lenses. Since the illuminated part of the mirror surfaces should not deviate from an ideal plane by more than $\lambda/10$ in order to obtain good interferograms, smaller beam diameters are advantageous.

The Mach–Zehnder interferometer has found a wide range of applications. Density variations in laminar or turbulent gas flows can be seen with this technique and the optical quality of mirror substrates or interferometer plates can be tested with high sensitivity [143, 144].

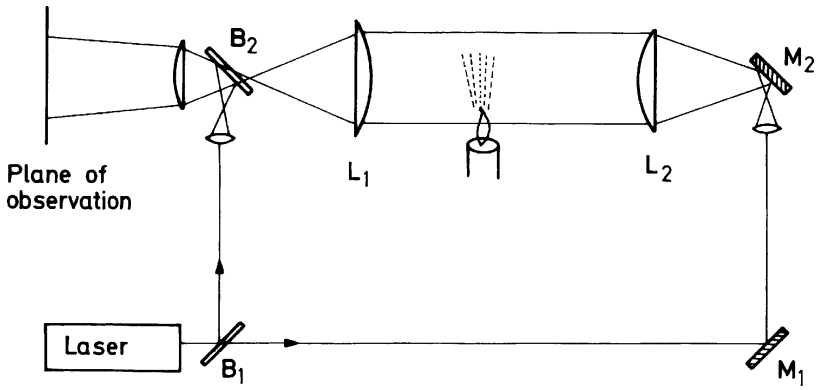

Figure 4.31 Laser interferometer for sensitive measurements of local variations of the index of refraction in extended samples, for example, in air above a candle flame

In order to get quantitative information of the local variation of the optical path through the sample, it is useful to generate a fringe pattern for calibration purposes by slightly tilting the plates B_1, M_1 and B_2, M_2 in Fig. 4.31, which makes the interferometer slightly asymmetric. Assume that B_1 and M_1 are tilted clockwise around the z-direction by a small angle β and the pair B_2, M_2 is tilted counterclockwise by the same angle β. The optical path between B_1 and M_1 is then $\Delta_1 = 2a\cos(\alpha+\beta)$, whereas $B_2M_2 = \Delta_2 = 2a\cos(\alpha-\beta)$. After being recombined, the two beams therefore have the path difference

$$\Delta = \Delta_2 - \Delta_1 = 2a[\cos(\alpha-\beta) - \cos(\alpha+\beta)] = 4a\sin\alpha\sin\beta\ , \qquad (4.43)$$

which depends on the angle of incidence α. In the plane of observation, an interference pattern of parallel fringes with path differences $\Delta = m\cdot\lambda$ is observed with an angular separation $\Delta\epsilon$ between the fringes m and $m+1$ given by $\Delta\epsilon = \alpha_m - \alpha_{m+1} = \lambda/(4a\sin\beta\cos\alpha)$.

A sample in path 3 introduces an additional path difference

$$\Delta s(\beta) = (n-1)L/\cos\beta$$

depending on the local refractive index n and the path length L through the sample. The resulting phase difference shifts the interference pattern by an angle $\gamma = (n-1)(L/\lambda)\Delta\varepsilon$. Using a lens with a focal length f, which images the interference pattern onto the plane O, gives the spatial distance $\Delta y = f\Delta\varepsilon$ between neighboring fringes. The additional path difference caused by the sample shifts the interference pattern by $N = (n-1)(L/\lambda)$ fringes.

Figure 4.32 shows for illustration the interferogram of the convection zone of hot air above a candle flame, placed below one arm of the laser interferometer in Fig. 4.31. It can be seen that the optical path through this zone changes by many wavelengths.

The Mach–Zehnder interferometer has been used in spectroscopy to measure the refractive index of atomic vapors in the vicinity of spectral lines (Sect. 3.1). The

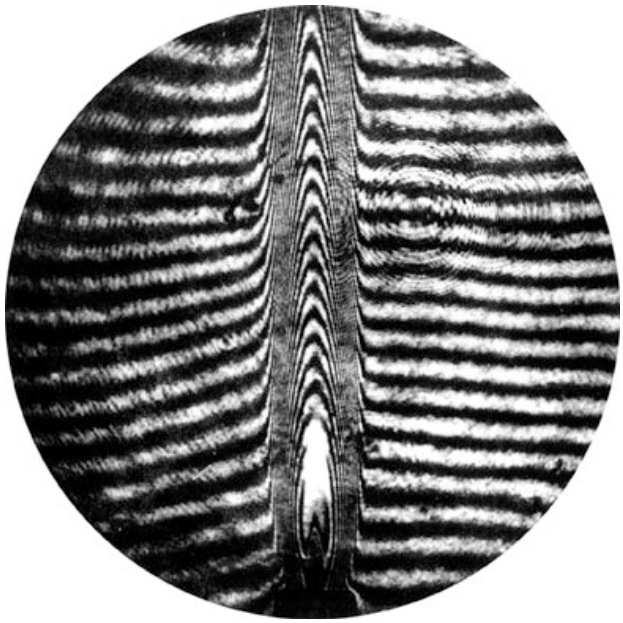

Figure 4.32 Interferogram of the density profile in the convection zone above a candle flame (H. Rottenkolber) [143]

experimental arrangement (Fig. 4.33) consists of a combination of a spectrograph and an interferometer, where the plates B_1, M_1 and B_2, M_2 are tilted in such a direction that without the sample the parallel interference fringes with the separation $\Delta y(\lambda) = f\Delta\varepsilon$ are perpendicular to the entrance slit, which is parallel to the y-direction. The spectrograph disperses the fringes with different wavelengths λ_i in the z-direction. Because of the wavelength-dependent refractive index $n(\lambda)$ of the atomic vapor (Sect. 3.1.3), the fringe shift follows a dispersion curve in the vicinity of the spectral line (Fig. 4.34). The dispersed fringes look like hooks around an absorption line, which gave this technique the name *hook method*. To compensate for background shifts caused by the windows of the absorption cell, a compensating plate is inserted into the second arm. This technique was developed in 1912 by Rozhdestvenski [146] in St. Petersburg. For more details of the Hook method, see [144–146].

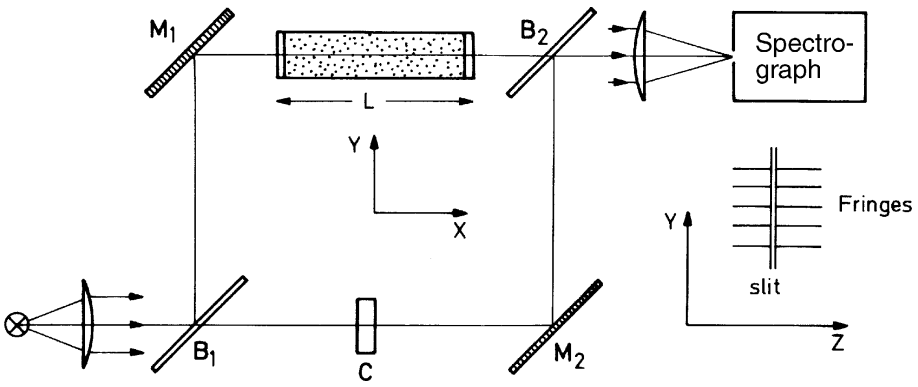

Figure 4.33 Combination of Mach–Zehnder interferometer and spectrograph used for the hook method

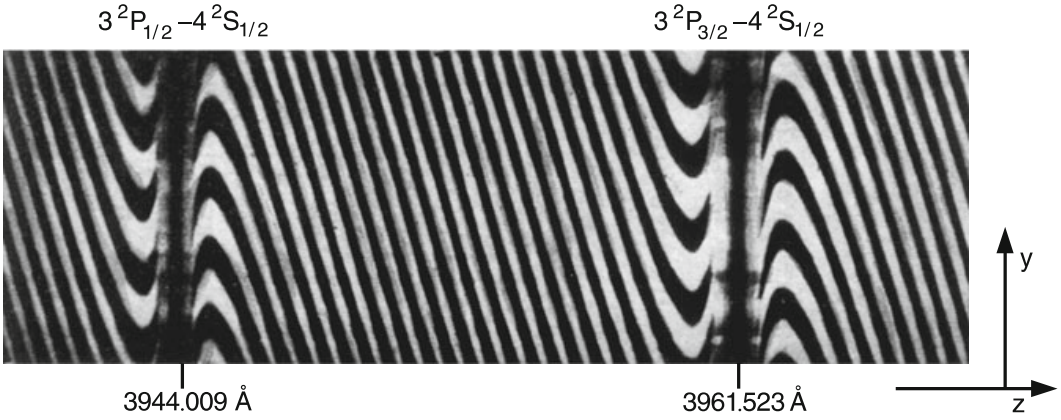

Figure 4.34 Position of fringes as a function of wavelength around the absorption line doublet of aluminium atoms, as observed behind the spectrograph [145]

4.2.5 Sagnac Interferometer

In the Sagnac interferometer (Fig. 4.35), the beam splitter BS splits the incoming beam into a transmitted beam and a reflected beam. The two beams circulate in opposite directions in the x, y-plane through the ring interferometer. If the whole interferometer rotates clockwise around an axis in the z-direction through the center of the x–y area around which the beams circulate, the optical path for the clockwise-circulating beam becomes longer than that for the counterclockwise running beam (the Sagnac effect). This causes a phase difference between the two beams and the intensity of the interfering beams as measured in the observation plane changes with the angular speed of rotation Ω. The phase shift between the two partial waves is

$$\Delta\phi = 8\pi A \cdot \boldsymbol{n} \cdot \boldsymbol{\Omega} / (\lambda \cdot c) \tag{4.44}$$

where A is the area inside the circulating beams, $\boldsymbol{n}$ is the unit vector perpendicular to the area A, λ the wavelength of the optical waves, and c the velocity of light. Using such a device angular velocities of less than $0.1°/\mathrm{h}$ (5×10^{-7} rad/s) can be detected. Using optical fibers the optical beams can circulate N times ($N = 100$–$10{,}000$) around the area A, and the effective area in (4.44) becomes $N \cdot A$, which increases the sensitivity considerably.

Such a device with three orthogonal Sagnac interferometers can be used as a navigation system, because the Earth's rotation causes a Sagnac effect that depends on the angle between the surface normal $\boldsymbol{n}$ and the Earth's axis of rotation $\boldsymbol{\omega}$; i.e., on the geographical latitude.

The Sagnac effect can be also explained by the Doppler effect: upon reflection at a mirror moving at a velocity v, the frequency ν of the reflected beam is shifted by $\Delta\nu = 2\nu \cdot v/c$. The frequencies of the two waves circulating in opposite directions

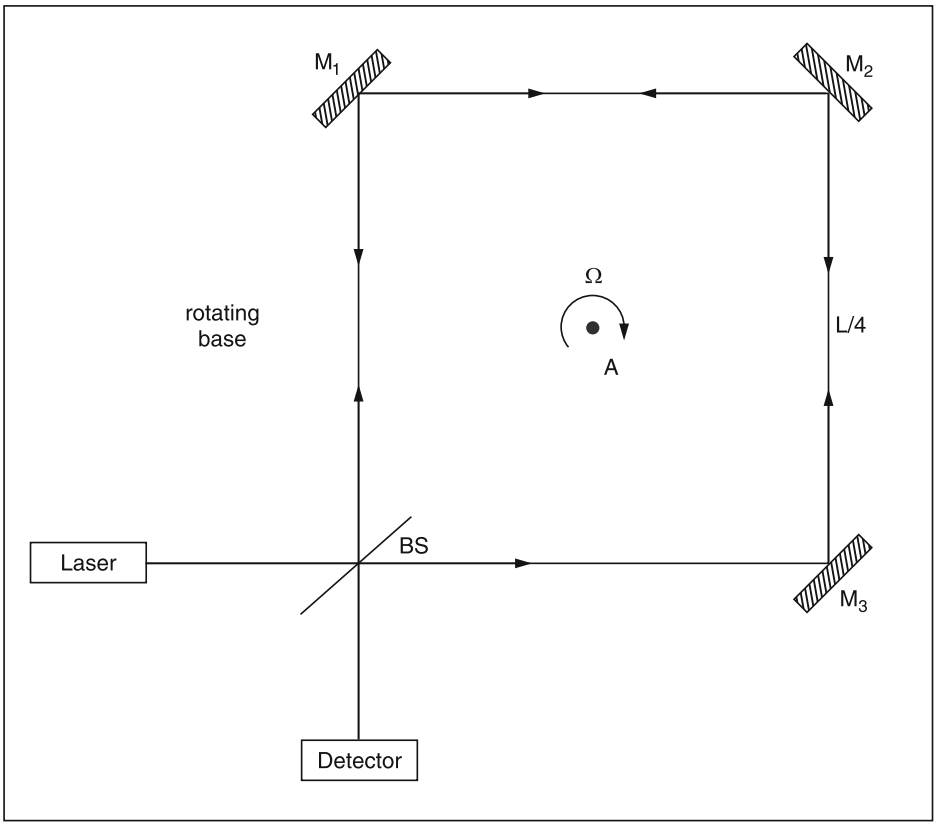

Figure 4.35 Sagnac interferometer

are therefore shifted away from each other by

$$\Delta\nu = 4A/(L \cdot \lambda)\boldsymbol{n} \cdot \boldsymbol{\Omega} \tag{4.45}$$

where L is the path length for one round trip in the ring interferometer. Since $\Delta\phi = (2\pi L/c)\Delta\nu$, both equations are equivalent, although the detection technique is different. The determination of the phase shift is based on measuring the intensity change at the detector, while the beat frequency $\Delta\nu$ can be directly counted with high precision [147].

4.2.6 Multiple-Beam Interference

In a grating spectrometer, the interfering partial waves emitted from the different grooves of the grating all have the same amplitude. In contrast, in multiple-beam interferometers these partial waves are produced by multiple reflection at plane or curved surfaces and their amplitude decreases with increasing number of reflections. The resultant total intensity therefore differs from (4.25).

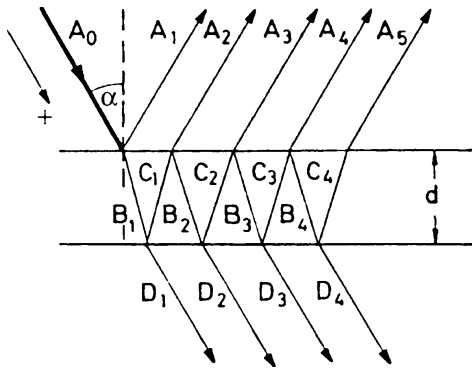

Figure 4.36 Multiple-beam interference at two plane-parallel partially reflecting surfaces

a) Transmitted and Reflected Intensity

Assume that a plane wave $E = A_0 \exp[\mathrm{i}(\omega t - kx)]$ is incident at the angle α on a plane transparent plate with two parallel, partially reflecting surfaces (Fig. 4.36). At each surface the amplitude A_i is split into a reflected component $A_\mathrm{R} = A_i\sqrt{R}$ and a refracted component $A_\mathrm{T} = A_i\sqrt{1-R}$, neglecting absorption. The reflectivity $R = I_\mathrm{R}/I_i$ depends on the angle of incidence α and on the polarization of the incident wave. Provided the refractive index n is known, R can be calculated from Fresnel's formulas [116]. From Fig. 4.36, the following relations are obtained for the amplitudes A_i of waves reflected at the upper surface, B_i of refracted waves, C_i of waves reflected at the lower surface, and D_i of transmitted waves

$$
\begin{aligned}
&|A_1| = \sqrt{R}\,|A_0|\,, && |B_1| = \sqrt{1-R}\,|A_0|\,,\\
&|C_1| = \sqrt{R(1-R)}\,|A_0|\,, && |D_1| = (1-R)\,|A_0|\,,\\
&|A_2| = \sqrt{1-R}\,|C_1| = (1-R)\sqrt{R}\,|A_0|\,, && |B_2| = R\sqrt{1-R}\,|A_0|\,,\\
&|C_2| = R\sqrt{R(1-R)}\,|A_0|\,, && |D_2| = R(1-R)\,|A_0|\,,\\
&|A_3| = \sqrt{1-R}\,|C_2| = R^{3/2}(1-R)\,|A_0|\,, && \ldots\,.
\end{aligned}
\tag{4.46}
$$

This scheme can be generalized to the equations

$$|A_{i+1}| = R|A_i|\,, \quad i \geq 2\,, \tag{4.47a}$$

$$|D_{i+1}| = R|D_i|\,, \quad i \geq 1\,. \tag{4.47b}$$

Two successively reflected partial waves E_i and E_{i+1} have the optical path difference (Fig. 4.37)

$$\Delta s = (2nd/\cos\beta) - 2d\tan\beta\sin\alpha\,.$$

Because $\sin\alpha = n\sin\beta$, this can be reduced to

$$\Delta s = 2nd\cos\beta = 2nd\sqrt{1-\sin^2\beta} = 2d\sqrt{n^2-\sin^2\alpha}\,, \tag{4.48a}$$

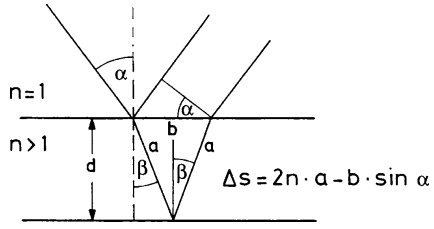

Figure 4.37 Optical path difference between two beams being reflected from the two surfaces of a plane-parallel plate

if the refractive index within the plane-parallel plate is $n > 1$ and outside the plate $n = 1$. This path difference causes a corresponding phase difference

$$\phi = 2\pi\Delta s/\lambda + \Delta\phi \ , \tag{4.48b}$$

where $\Delta\phi$ takes into account possible phase changes caused by the reflections. For instance, the incident wave with amplitude A_1 suffers the phase jump $\Delta\phi = \pi$ while being reflected at the medium with $n > 1$. Including this phase jump, we can write

$$A_1 = \sqrt{R}A_0 \exp(\mathrm{i}\pi) = -\sqrt{R}A_0 \ .$$

The total amplitude A of the reflected wave is obtained by summation over all partial amplitudes A_i, taking into account the different phase shifts,

$$\begin{aligned} A &= \sum_{m=1}^{p} A_m \mathrm{e}^{\mathrm{i}(m-1)\phi} = -\sqrt{R}A_0 + \sqrt{R}A_0(1-R)\mathrm{e}^{\mathrm{i}\phi} + \sum_{m=3}^{p} A_m \mathrm{e}^{\mathrm{i}(m-1)\phi} \\ &= -\sqrt{R}A_0 \left[1 - (1-R)\mathrm{e}^{\mathrm{i}\phi} \sum_{m=0}^{p-2} R^m \mathrm{e}^{\mathrm{i}m\phi} \right] \ . \end{aligned} \tag{4.49}$$

For vertical incidence ($\alpha = 0$), or for an infinitely extended plate, we have an infinite number of reflections. The geometrical series in (4.49) has the limit $(1 - R\mathrm{e}^{\mathrm{i}\phi})^{-1}$ for $p \to \infty$. We obtain for the total amplitude

$$A = -\sqrt{R}A_0 \frac{1 - \mathrm{e}^{\mathrm{i}\phi}}{1 - R\mathrm{e}^{\mathrm{i}\phi}} \ . \tag{4.50}$$

The intensity $I = 2c\epsilon_0 AA^*$ of the reflected wave is then, with $I_0 = 2c\varepsilon_0 A_0 A_0^*$,

$$\boxed{I_\mathrm{R} = I_0 R \frac{4\sin^2(\phi/2)}{(1-R)^2 + 4R\sin^2(\phi/2)} \ .} \tag{4.51a}$$

In an analogous way, we find for the total transmitted amplitude

$$D = \sum_{m=1}^{\infty} D_m \mathrm{e}^{\mathrm{i}(m-1)\phi} = (1-R)A_0 \sum_{0}^{\infty} R^m \mathrm{e}^{\mathrm{i}m\phi} \ ,$$

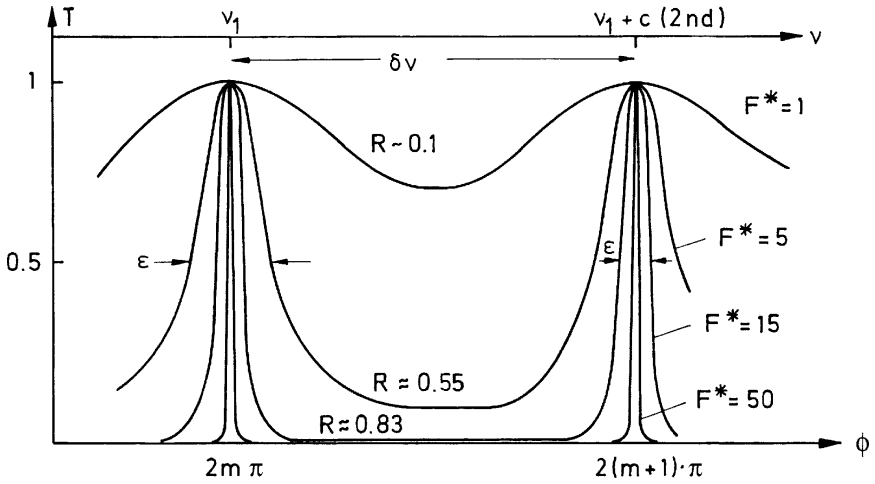

Figure 4.38 Transmittance of an absorption-free multiple-beam interferometer as a function of the phase difference ϕ for different values of the finesse F^*

which gives the total transmitted intensity

$$\boxed{I_\mathrm{T} = I_0 \frac{(1-R)^2}{(1-R)^2 + 4R\sin^2(\phi/2)}\,.} \tag{4.52a}$$

Equations (4.51a, 4.52a) are called the *Airy formulas*. Since we have neglected absorption, we should have $I_\mathrm{R} + I_\mathrm{T} = I_0$, as can easily be verified from (4.51a, 4.52a).

The abbreviation $F = 4R/(1-R)^2$ is often used, which allows the Airy equations to be written in the form

$$\boxed{I_\mathrm{R} = I_0 \frac{F\sin^2(\phi/2)}{1 + F\sin^2(\phi/2)}\,,} \tag{4.51b}$$

$$\boxed{I_\mathrm{T} = I_0 \frac{1}{1 + F\sin^2(\phi/2)}\,.} \tag{4.52b}$$

Figure 4.38 illustrates (4.52b) for different values of the reflectivity R. The maximum transmittance is $T = 1$ for $\phi = 2m\pi$. At these maxima $I_\mathrm{T} = I_0$, therefore the reflected intensity I_R is zero. The minimum transmittance is

$$T^{\min} = \frac{1}{1+F} = \left(\frac{1-R}{1+R}\right)^2\,.$$

Example: For $R = 0.98 \Rightarrow T^{\min} = 10^{-4}$.
For $R = 0.90 \Rightarrow T^{\min} = 2.8 \times 10^{-3}$.

The ratio

$$C = I_{T_{\max}}/I_{T_{\min}} = 1 + F = \left(\frac{1+R}{1-R}\right)^2$$

is named the *contrast* of the interferometer.

b) Free Spectral Range and Finesse

The frequency range $\delta\nu$ between two maxima is the *free spectral range of the interferometer*. Interference maxima occur for $\Delta s = m \cdot \lambda$ ($m = 1, 2, 3, \ldots$) which corresponds to a phase difference $\Delta\Phi = m \cdot 2\pi$. The free spectral range $\delta\lambda$ is obtained from

$$\Phi_1 - \Phi_2 = 2\pi\Delta s/\lambda_1 - 2\pi\Delta s/\lambda_2 = 2(m+1)\pi - 2m\pi = 2\pi$$
$$\rightarrow \Delta s \cdot (\lambda_2 - \lambda_1) = \lambda_1 \cdot \lambda_2 \approx \lambda^2 \Rightarrow \delta\lambda = \lambda^2/\Delta s \; .$$

Wegen $\nu = c/\lambda$ folgt

$$\delta\nu = -\left(c/\lambda^2\right)\,\delta\lambda = c/\Delta s = c/\left(2d\sqrt{n^2 - \sin^2\alpha}\right) \; . \tag{4.53a}$$

For vertical incidence ($\alpha = 0$), the free spectral range becomes

$$\boxed{|\delta\nu|_{\alpha=0} = \frac{c}{2nd} \; .} \tag{4.53b}$$

The full halfwidth $\epsilon = |\phi_1 - \phi_2|$ with $I(\phi_1) = I(\phi_2) = I_0/2$ of the transmission maxima in Fig. 4.38 expressed in phase differences is calculated from (4.52a), (4.52b) as

$$\epsilon = 4\arcsin\left(\frac{1-R}{2\sqrt{R}}\right) \; , \tag{4.54a}$$

which reduces for $R \approx 1 \Rightarrow (1-R) \ll R$ to

$$\epsilon = \frac{2(1-R)}{\sqrt{R}} = \frac{4}{\sqrt{F}} \; . \tag{4.54b}$$

In frequency units, the free spectral range $\delta\nu$ corresponds to a phase difference $\delta\phi = 2\pi$. Therefore the halfwidth $\Delta\nu$ becomes

$$\Delta\nu = \frac{\epsilon}{2\pi}\delta\nu \simeq \frac{2\delta\nu}{\pi\sqrt{F}}$$

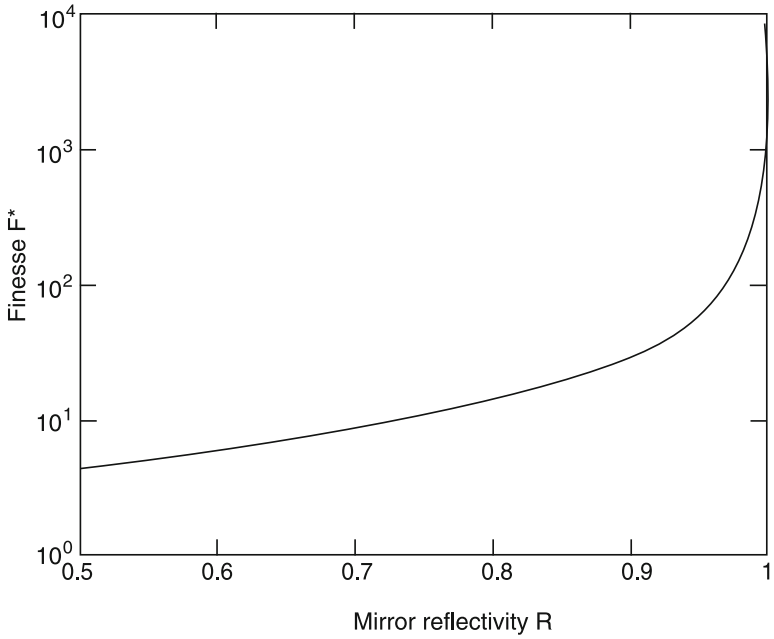

Figure 4.39 Finesse F_R^* of a Fabry–Perot interferometer as a function of the mirror reflectivity R

which yields for vertical incidence with (4.53b)

$$\Delta\nu = \frac{c}{2nd}\frac{1-R}{\pi\sqrt{R}} \,. \tag{4.54c}$$

The ratio $\delta\nu/\Delta\nu$ of free spectral range $\delta\nu$ to the halfwidth $\Delta\nu$ of the transmission maxima is called the *finesse* F^* of the interferometer. From (4.53b) and (4.54c) we obtain for the "reflectivity finesse" F_R^*

$$\boxed{F_R^* = \frac{\delta\nu}{\Delta\nu} = \frac{\pi\sqrt{R}}{1-R} = \frac{\pi}{2}\sqrt{F} \,.} \tag{4.55a}$$

The full halfwidth of the transmission peaks is then

$$\Delta\nu = \frac{\delta\nu}{F_R^*} \,. \tag{4.55b}$$

The finesse is a measure for the effective number of interfering partial waves in the interferometer. This means that for vertical incidence the maximum path difference between interfering waves is $\Delta s_{max} = F^* 2nd$. Figure 4.39 shows the finesse F_R^* as a function of the mirror reflectivity.

Since we have assumed an ideal plane-parallel plate with a perfect surface quality, the finesse (4.55a) is determined only by the reflectivity R of the surfaces. In practice, however, deviations of the surfaces from an ideal plane and slight inclinations of the two surfaces cause imperfect superposition of the interfering waves.

This results in a decrease and a broadening of the transmission maxima, which decreases the total finesse. If, for instance, a reflecting surface deviates by the amount λ/q from an ideal plane, the finesse cannot be larger than q. One can define the total finesse F^* of an interferometer by

$$\frac{1}{F^{*2}} = \sum_i \frac{1}{F_i^{*2}} \,, \tag{4.55c}$$

where the different terms F_i^* give the contributions to the decrease of the finesse caused by the different imperfections of the interferometer.

If, for instance, the surface of the mirror shows a parabolic deviation from a plane surface, i.e.,

$$S(r, \varphi) = S_0 + \alpha r^2$$

the finesse becomes (with $k = 2\pi/\lambda$ [148])

$$F^* = \frac{\pi}{[(1-R)^2/R + k^2\alpha^2]^{1/2}} \tag{4.56}$$

which yields

$$\frac{1}{F^{*2}} = \frac{(1-R)^2}{\pi^2 R} + \frac{4\alpha^2}{\lambda^2} = \frac{1}{F_{\mathrm{R}}^{*2}} + \frac{1}{F_f^{*2}} \tag{4.57}$$

where F_p is the finesse determined by the curvature of the mirror surface.

Example 4.10

A plane, nearly parallel plate has a diameter $D = 5\,\mathrm{cm}$, a thickness $d = 1\,\mathrm{cm}$, and a wedge angle of $0.2''$. The two reflecting surfaces have the reflectivity $R = 95\,\%$. The surfaces are flat to within $\lambda/50$, which means that no point of the surface deviates from an ideal plane by more than $\lambda/50$. The different contributions to the finesse are:

- Reflectivity finesse: $F_{\mathrm{R}}^* = \pi\sqrt{R}/(1-R) \simeq 60$;
- Surface finesse: $F_{\mathrm{S}} \simeq 50$;
- Wedge finesse: with a wedge angle of $0.2''$ the optical path between the two reflecting surfaces changes by about $0.1\lambda(\lambda = 0.5\,\mu\mathrm{m})$ across the diameter of the plate. For a monochromatic incident wave this causes imperfect interference and broadens the maxima corresponding to a finesse of about 20.

The total finesse is then $F^{*2} = 1/(1/60^2 + 1/50^2 + 1/20^2) \rightarrow F^* \simeq 17.7$.

This illustrates that high-quality optical surfaces are necessary to obtain a high total finesse [148]. It makes no sense to increase the reflectivity without

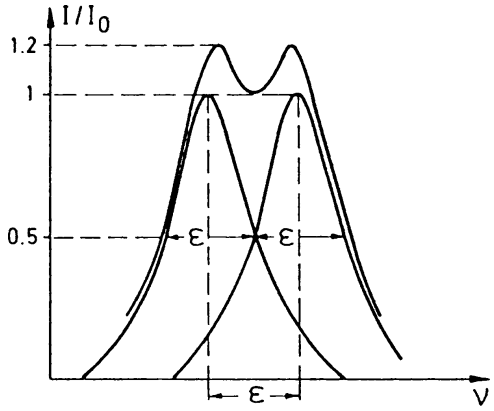

Figure 4.40 Transmitted intensity $I_T(\nu)$ for two closely spaced spectral lines at the limit of spectral resolution where the linespacing equals the halfwidth of the lines

a corresponding increase of the surface finesse. In our example the imperfect parallelism was the main cause for the low finesse. Decreasing the wedge angle to $0.1''$ increases the wedge finesse to 40 and the total finesse to 27.7.

A much larger finesse can be achieved using spherical mirrors, because the demand for parallelism is dropped. With sufficiently accurate alignment and high reflectivities, values of $F^* > 50{,}000$ are possible (Sect. 4.2.10).

c) Spectral Resolution

The spectral resolution, $\nu/\Delta\nu$ or $\lambda/\Delta\lambda$, of an interferometer is determined by the free spectral range $\delta\nu$ and by the finesse F^*. Two incident waves with frequencies ν_1 and $\nu_2 = \nu_1 + \Delta\nu$ can still be resolved if their frequency separation $\Delta\nu$ is larger than $\delta\nu/F^*$, which means that their peak separation should be larger than their full halfwidth.

Quantitatively this can be seen as follows: assume the incident radiation consists of two components with the intensity profiles $I_1(\nu - \nu_1)$ and $I_2(\nu - \nu_2)$ and equal peak intensities $I_1(\nu_1) = I_2(\nu_2) = I_0$. For a peak separation $\nu_2 - \nu_1 = \delta\nu/F^* = 2\delta\nu/\pi\sqrt{F}$, the total transmitted intensity $I(\nu) = I_1(\nu) + I_2(\nu)$ is obtained from (4.52a) as

$$I(\nu) = I_0 \left(\frac{1}{1 + F \sin^2(\pi\nu/\delta\nu)} + \frac{1}{1 + F \sin^2[\pi(\nu + \delta\nu/F^*)/\delta\nu]} \right) , \quad (4.58)$$

where the phase shift $\phi = 2\pi\Delta s/\lambda = 2\pi\Delta s(\nu/c) = 2\pi\nu/\delta\nu$ in (4.52b) has been expressed by the free spectral range $\delta\nu = c/2nd = c/\Delta s$, where Δs is the optical path difference between two successive partial waves in Fig. 4.36 for $\alpha = 0$. The function $I(\nu)$ is plotted in Fig. 4.40 around the frequency $\nu = (\nu_1 + \nu_2)/2$. For $\nu = \nu_1 = mc/2nd$, the first term in (4.58) becomes 1 and the second term can be derived with $\sin[\pi(\nu_1 + \delta\nu/F^*)/\delta\nu] = \sin\pi/F^* \simeq \pi/F^*$ and $F(\pi/F^*)^2 = 4$

to become 0.2. Inserting this into (4.58) yields $I(\nu = \nu_1) = 1.2I_0$, $I(\nu = (\nu_1 + \nu_2)/2) \simeq I_0$, and $I(\nu = \nu_2) = 1.2I_0$. This just corresponds to the Rayleigh criterion for the resolution of two spectral lines. The spectral resolving power of the interferometer is therefore

$$\boxed{\nu/\Delta\nu = (\nu/\delta\nu)F^* \rightarrow \Delta\nu = \delta\nu/F^* \,.} \tag{4.59}$$

This can be also expressed by the optical path differences Δs between two successive partial waves

$$\boxed{\frac{\nu}{\Delta\nu} = \frac{\lambda}{\Delta\lambda} = F^* \frac{\Delta s}{\lambda} \,.} \tag{4.60}$$

The resolving power of an interferometer is the product of finesse F^* and optical path difference $\Delta s/\lambda$ in units of the wavelength λ.

A comparison with the resolving power $\nu/\Delta\nu = mN = N\Delta s/\lambda$ of a grating spectrometer with N grooves shows that the finesse F^* can indeed be regarded as the effective number of interfering partial waves and $F^*\Delta s$ can be regarded as the maximum path difference between these waves.

The spectral resolution $\nu/\Delta\nu = \lambda/\Delta\lambda$ equals the maximum path-difference $\Delta s_{\max}/\lambda$ in units of the wavelength λ.

Example 4.11
$d = 1\,\text{cm}$, $n = 1.5$, $R = 0.98$, $\lambda = 500\,\text{nm}$. An interferometer with negligible wedge and high-quality surfaces, where the finesse is mainly determined by the reflectivity, achieves with $F^* = \pi\sqrt{R}/(1-R) = 155$ a resolving power of $\lambda/\Delta\lambda = 10^7$. This means that the instrument's linewidth is about $\Delta\lambda \sim 5 \times 10^{-5}\,\text{nm}$ or, in frequency units, $\Delta\nu = 60\,\text{MHz}$.

d) Influence of Absorption Losses

Taking into account the absorption $A = (1 - R - T)$ of each reflective surface, the transmitted intensity (4.52a), (4.52b) must be modified to

$$I_\text{T} = I_0 \frac{T^2}{(A+T)^2} \frac{1}{[1 + F\sin^2(\delta/2)]} \,, \tag{4.61a}$$

where $T^2 = T_1 T_2$ is the product of the transmittance of the two reflecting surfaces. The absorption causes three effects:

a) The maximum transmittance is decreased by the factor

$$\frac{I_\mathrm{T}}{I_0} = \frac{T^2}{(A+T)^2} = \frac{T^2}{(1-R)^2} < 1\ . \tag{4.61b}$$

Note that even a small absorption of each reflecting surface results in a drastic reduction of the total transmittance. For $A = 0.05$, $R = 0.9 \rightarrow T = 0.05 \Rightarrow T^2/(1-R)^2 = 0.25$. This illustrates, that even for a small absorption of $A = 5\,\%$ the transmitted intensity drops to 25 % of the absorption-free transmittance.

b) For a given transmission factor T, the reflectivity $R = 1-A-T$ decreases with increasing absorption. The quantity

$$F = \frac{4R}{(1-R)^2} = \frac{4(1-T-A)}{(T+A)^2} \tag{4.61c}$$

decreases with increasing A. For the example above we obtain $F = 360$. This makes the transmission peaks broader because of the decreasing number of interfering partial waves. The *contrast*

$$C = \frac{I_\mathrm{T}^\mathrm{max}}{I_\mathrm{T}^\mathrm{min}} = 1 + F = \left(\frac{1+R}{1-R}\right)^2 \tag{4.61d}$$

of the transmitted intensity also decreases. If absorption is taken into account, we can insert $R = 1 - A - T$ and obtain for the contrast

$$C = I_{T_\mathrm{max}}/I_{T_\mathrm{min}} = (2-A-T)^2/(A+T)^2\ .$$

For $A > 0$ but constant transmission T the reflectivity R and therefore the contrast C decreases.

Example 4.12
$R = 0.95$, $A = 0.03$, $T = 0.02 \Rightarrow C = 1521$.
Without absorption is $A = 0$, $T = 0.02$ and $R = 0.98 \Rightarrow C = 9801$.

c) The absorption causes a phase shift $\Delta\phi$ at each reflection, which depends on the wavelength λ, the polarization, and the angle of incidence α [116]. This effect causes a wavelength-dependent *shift* of the maxima.

4.2.7 Plane Fabry–Perot Interferometer

A practical realization of the multiple beam-interference discussed in this section may use either a solid plane-parallel glass or fused quartz plate with two coated reflecting surfaces (Fabry–Perot etalon, Fig. 4.41a) or two separate plates, where one

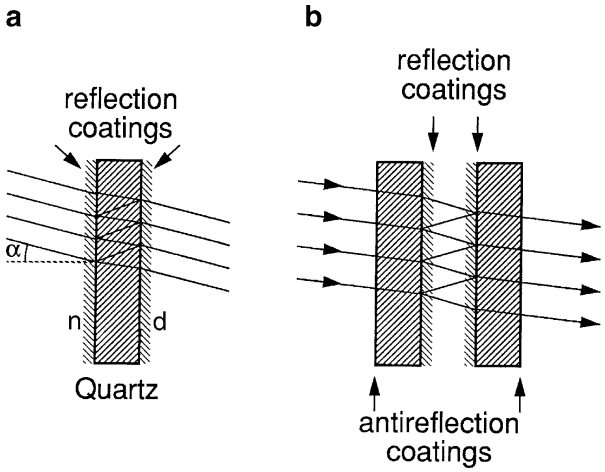

Figure 4.41 Two realizations of a Fabry–Perot interferometer: **a** solid etalon; **b** air-spaced plane-parallel reflecting surfaces

surface of each plate is coated with a reflection layer. The two reflecting surfaces are opposed and are aligned to be as parallel as achievable (Fabry–Perot interferometer (FPI), Fig. 4.41b). The outer surfaces are coated with antireflection layers in order to avoid reflections from these surfaces that might overlap the interference pattern. Furthermore, they have a slight angle against the inner surfaces (wedge).

Both devices can be used for parallel as well as for divergent incident light. We now discuss them in more detail, first considering their illumination with *parallel* light.

a) The Plane FPI as a Transmission Filter

In laser spectroscopy, etalons are mainly used as wavelength-selective transmission filters within the laser resonator to narrow the laser bandwidth (Sect. 5.4). The wavelength λ_m or frequency ν_m for the transmission maximum of mth order, where the optical path between successive beams is $\Delta s = m\lambda$, can be deduced from (4.48a) and Fig. 4.37 to be

$$\lambda_m = \frac{2d}{m}\sqrt{n^2 - \sin^2\alpha} = \frac{2nd}{m}\cos\beta \ , \tag{4.62a}$$

$$\nu_m = \frac{mc}{2nd\cos\beta} \ . \tag{4.62b}$$

For all wavelengths $\lambda = \lambda_m$ ($m = 1, 2, \ldots$) in the incident light, the phase difference between the transmitted partial waves becomes $\delta = 2m\pi$ and the transmitted intensity is, according to (4.61a)–(4.61d),

$$I_\mathrm{T} = \frac{T^2}{(1-R)^2} I_0 = \frac{T^2}{(A+T)^2} I_0 \ , \tag{4.63}$$

where $A = 1 - T - R$ is the absorption of the etalon (substrate absorption plus absorption of one reflecting surface). The reflected waves interfere destructively for

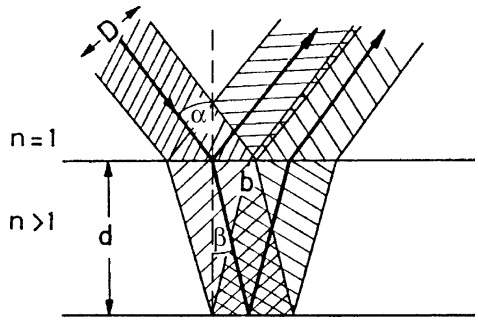

Figure 4.42 Incomplete interference of two reflected beams with finite diameter D, causing a decrease of the maximum transmitted intensity

$\lambda = \lambda_m$ and the *reflected* intensity becomes zero for $A = 0$ while the *transmitted* intensity is $I_\mathrm{T} = I_0$.

For $A > 0$ there remains a small residual reflected intensity and $I_\mathrm{T}/I_0 < 1$.

Note, however, that this is only true for $A \ll 1$ and infinitely extended plane waves, where the different reflected partial waves completely overlap. If the incident wave is a laser beam with the finite diameter D, the different reflected partial beams do *not* completely overlap because they are laterally shifted by $\Delta = b \cos\alpha$ with $b = 2d \tan\beta$ (Fig. 4.42). For a rectangular intensity profile of the laser beam, the fraction Δ/D of the reflected partial amplitudes does not overlap and cannot interfere destructively. This means that, even for maximum transmission, the reflected intensity is not zero but a background reflection remains, which is missing in the transmitted light. For small angles α, one obtains for the intensity loss per transit due to reflection [149] for a rectangular beam profile

$$\frac{I_\mathrm{R}}{I_0} = \frac{4R}{(1-R)^2}\left(\frac{2\alpha d}{nD}\right)^2 . \tag{4.64a}$$

For a Gaussian beam profile the calculation is more difficult, and the solution can only be obtained numerically. The result for a Gaussian beam with the radius w (Sect. 5.3) is [150]

$$\frac{I_\mathrm{R}}{I_0} \simeq \frac{8R}{(1-R)^2}\left(\frac{2d\alpha}{nw}\right)^2 . \tag{4.64b}$$

A parallel light beam with the diameter D passing a plane-parallel plate with the angle of incidence α therefore suffers reflection losses in addition to the eventual absorption losses. The reflection losses increase with α^2 and are proportional to the ratio $(d/D)^2$ of the etalon thickness d and the beam diameter D (walk-off losses).

Example 4.13
$d = 1\,\mathrm{cm}$, $D = 0.2\,\mathrm{cm}$, $n = 1.5$, $R = 0.3$, $\alpha = 1^\circ \mathrel{\hat{=}} 0.017\,\mathrm{rad} \rightarrow I_\mathrm{R}/I_0 = 0.05$, which means 5 % walk-off losses.

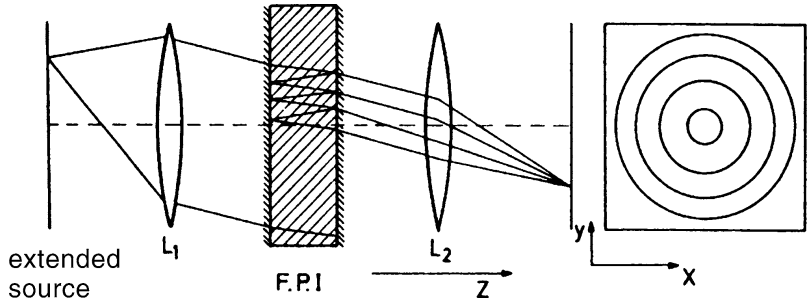

Figure 4.43 The interference ring system of the transmitted intensity may be regarded as wavelength-selective imaging of corresponding ring areas of an extended light source

The transmission peak λ_m of the etalon can be shifted by tilting the etalon. According to (4.62a), (4.62b) the wavelength λ_m *decreases* with increasing angle of incidence α. The walk-off losses, however, limit the tuning range of tilted etalons within a laser resonator. With increasing angle α, the losses may become intolerably large.

b) Illumination with Divergent Light

Illuminating the FPI with divergent monochromatic light (e.g., from an extended source or from a laser beam behind a diverging lens), a continuous range of incident angles α is offered to the FPI, which transmits, for a wavelength λ_m, those directions α_m that obey (4.62a). We then observe an interference pattern of bright rings in the transmitted light (Fig. 4.43). Since the reflected intensity $I_\mathrm{R} = I_0 - I_\mathrm{T}$ is complementary to the transmitted one, a corresponding system of dark rings appears in the reflected light at the same angles of incidence α_m.

When β is the angle of inclination to the interferometer axis inside the FPI, the transmitted intensity is maximum, according to (4.62a), (4.62b), for

$$m\lambda = 2nd\cos\beta \ , \tag{4.65}$$

where n is the refractive index between the reflecting planes. Let us number the rings by the integer p, beginning with $p = 0$ for the central ring. With $m = m_0 - p$, we can rewrite (4.65) for small angles β_p as

$$(m_0 - p)\lambda = 2nd\cos\beta_p \sim 2nd(1 - \beta_p^2/2) = 2nd\left[1 - \frac{1}{2}\left(\frac{n_0\alpha_p}{n}\right)^2\right] \ , \tag{4.66}$$

where n_0 is the refractive index of air, and Snell's law $\sin\alpha \simeq \alpha = (n/n_0)\beta$ has been used (Fig. 4.44).

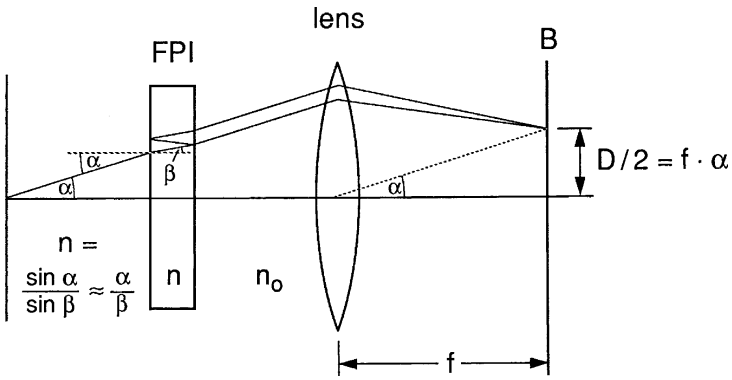

Figure 4.44 Illustration of (4.67a), (4.67b)

When the interference pattern is imaged by a lens with the focal length f into the plane of the photoplate, we obtain for the ring diameters $D_p = 2f\alpha_p$ the relations

$$(m_0 - p)\lambda = 2nd\left[1 - (n_0/n)^2 D_p^2/\left(8f^2\right)\right] , \tag{4.67a}$$

$$(m_0 - p - 1)\lambda = 2nd\left[1 - (n_0/n)^2 D_{p+1}^2/\left(8f^2\right)\right] . \tag{4.67b}$$

Subtracting the second equation from the first one yields

$$D_{p+1}^2 - D_p^2 = \frac{4nf^2}{n_0^2 d}\lambda . \tag{4.68}$$

For the smallest ring with $p = 0$, (4.66) becomes

$$m_0\lambda = 2nd\left(1 - \beta_0^2/2\right) \Rightarrow m_0\lambda + nd\beta_0^2 = 2nd , \tag{4.69}$$

which can be written as

$$(m_0 + \epsilon)\lambda = 2nd . \tag{4.70}$$

The "excess" $\epsilon < 1$, also called *fractional interference order*, can be obtained from a comparison of (4.69) and (4.70) as

$$\epsilon = nd\beta_0^2/\lambda = (n_0/n)d\alpha_0^2/\lambda . \tag{4.71}$$

Inserting ϵ from (4.70) into (4.67a) yields the relation

$$D_p^2 = \frac{8n^2 f^2}{n_0^2(m_0 + \epsilon)}(p + \epsilon) . \tag{4.72}$$

A linear fit of the squares D_p^2 of the measured ring diameters versus the ring number p yields the excess ϵ and therefore from (4.70) the wavelength λ, provided the refractive index n and the value of d of the plate separation are known from a previous calibration of the interferometer. However, the wavelength is determined by

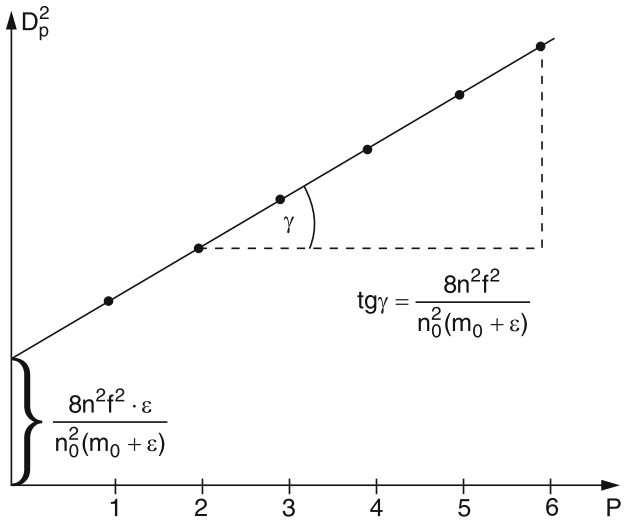

Figure 4.45 Determination of the access ϵ from the plot of D_p^2 versus p

(4.70) only modulo a free spectral range $\delta\lambda = \lambda^2/(2nd)$. This means that all wavelengths λ_m differing by m free spectral ranges produce the same ring systems. For an absolute determination of λ, the integer order m_0 must be known.

When illuminated with strictly parallel light parallel to the interferometer axis the plane of observation behind the FPI is completely dark for $\lambda \neq 2nd \cos\beta/m$ and uniformly illuminated for $\lambda = 2nd \cos\beta/m$. The reflected intensity is zero for $\lambda = 2nd \cos\beta/m$, otherwise the reflection coefficient is $R = 1$ which means that the total incident intensity is reflected back into the source.

Illumination with divergent light produces a ring system of bright rings with a dark background. In the reflected light a dark ring system on a bright background appears. If the light beam is divergent in the *x*-direction but parallel in the *y*-direction the ring system changes into a system of parallel straight bright or dark lines.

The experimental scheme for the absolute determination of λ utilizes a combination of FPI and spectrograph in a so-called *crossed arrangement* (Fig. 4.46), where the ring system of the FPI is imaged onto the entrance slit of a spectrograph. The spectrograph disperses the slit images $S(\lambda)$ with a medium dispersion in the x-direction (Sect. 4.1), the FPI provides high dispersion in the y-direction. The resolution of the spectrograph must only be sufficiently high to separate the images of two wavelengths differing by one free spectral range of the FPI. Figure 4.47 shows, for illustration, a section of the Na_2 fluorescence spectrum excited by an argon laser line. The ordinate corresponds to the FPI dispersion and the abscissa to the spectrograph dispersion [151].

The angular dispersion $d\beta/d\lambda$ of the FPI can be deduced from (4.66)

$$\frac{d\beta}{d\lambda} = \left(\frac{d\lambda}{d\beta}\right)^{-1} = m/(2nd \sin\beta) = \frac{1}{\lambda_m \sin\beta} \quad \text{with } \lambda_m = 2nd/m \,. \qquad (4.73)$$

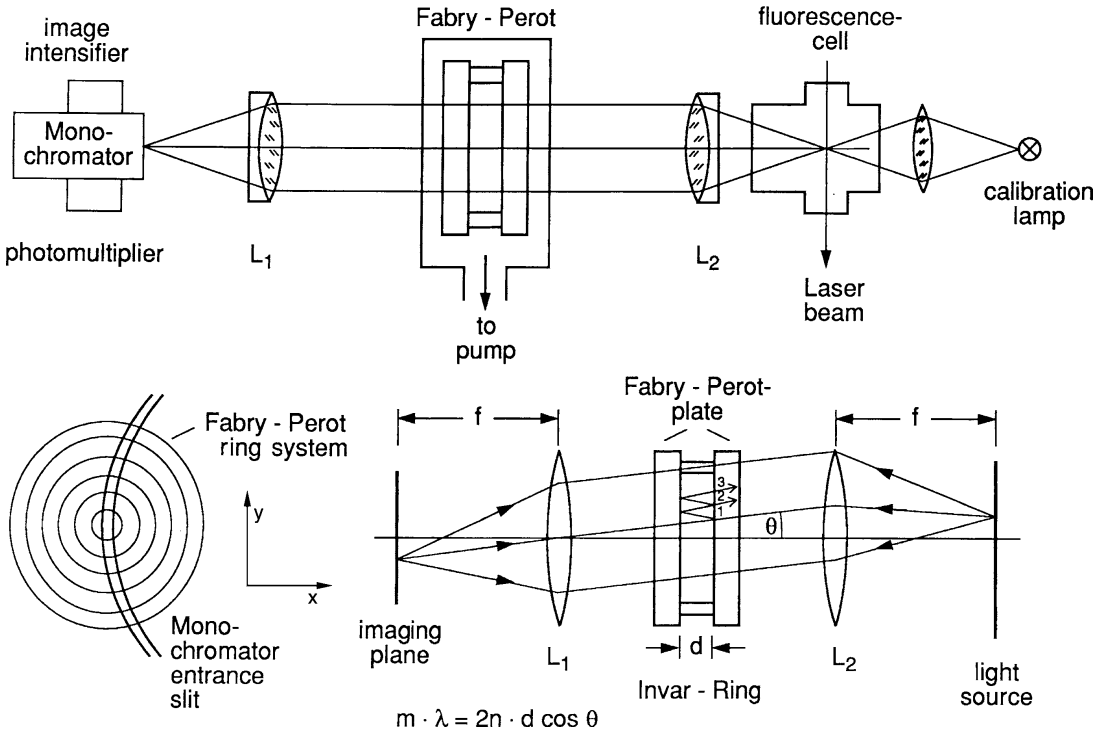

Figure 4.46 Combination of FPI and spectrograph for the unambiguous determination of the integral order m_0

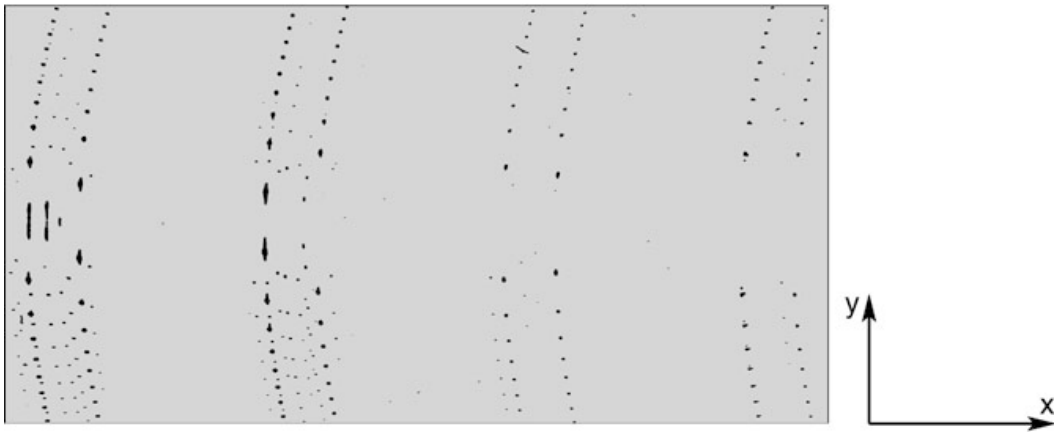

Figure 4.47 Section of the argon laser-excited fluorescence spectrum of Na_2 obtained with the arrangement of crossed FPI and spectrograph shown in Fig. 4.46 [151]

Equation (4.73) shows that the angular dispersion becomes infinite for $\beta \to 0$. The linear dispersion of the ring system on the photoplate is

$$\frac{\mathrm{d}D}{\mathrm{d}\lambda} = f\frac{\mathrm{d}\beta}{\mathrm{d}\lambda} = \frac{f}{\lambda_m \sin\beta} \,. \tag{4.74}$$

Example 4.14
$f = 50\,\text{cm}$, $\lambda = 0.5\,\mu\text{m}$. At a distance of 1 mm from the ring center is $\beta = 0.1/50$ and we obtain a linear dispersion of $\mathrm{d}D/\mathrm{d}\lambda = 500\,\text{mm/nm}$. This is at least one order of magnitude larger than the dispersion of a large spectrograph.

c) The Air-Spaced FPI

Different from the solid etalon, which is a plane-parallel plate coated on both sides with reflecting layers, the plane FPI consists of two wedged plates, each having one high-reflection and one antireflection coating (Fig. 4.41b). The finesse of the FPI critically depends, apart from the reflectivity R and the optical surface quality, on the parallel alignment of the two reflecting surfaces. The advantage of the air-spaced FPI, that any desired free spectral range can be realized by choosing the corresponding plate separation d, must be paid for by the inconvenience of careful alignment. Instead of changing the angle of incidence α, wavelength tuning can be also achieved for $\alpha = 0$ by variation of the optical path difference $\Delta s = 2nd$, either by changing d with piezoelectric tuning of the plate separation, or by altering the refractive index by a pressure change in the container enclosing the FPI.

The tunable FPI is used for high-resolution spectroscopy of line profiles. The transmitted intensity $I_\mathrm{T}(p)$ as a function of the optical path difference nd is given by the convolution

$$I_\mathrm{T}(\nu) = I_0(\nu)T(nd,\lambda)\ ,$$

where the transmission of the FPI $T(nd,\lambda) = T(\phi)$ can be obtained from (4.52a), (4.52b).

With photoelectric recording (Fig. 4.48), the large dispersion at the ring center can be utilized. The light source LS is imaged onto a small pinhole P1, which serves as a point source in the focal plane of L1. The parallel light beam passes the FPI, and the transmitted intensity is imaged by L2 onto another pinhole P2 in front of the detector. All light rays within the cone $\cos\beta_0 \leq m_0\lambda/(nd)$, where β is the angle against the interferometer axis, contribute according to (4.66) to the central fringe. If the optical path length nd is tuned, the different transmission orders with $m = m_0, m_0+1, m_0+2, \ldots$ are successively transmitted for a wavelength λ according to $m\lambda = 2nd$. Light sources that come close to being a point source, can be realized when a focused laser beam crosses a sample cell and the laser-induced fluorescence emitted from a small section of the beam length is imaged through the FPI onto the entrance slit of a monochromator, which is tuned to the desired wavelength interval $\Delta\lambda$ around λ_m (Fig. 4.46). If the spectral interval $\Delta\lambda$ resolved by the monochromator is smaller then the free spectral range $\delta\lambda$ of the

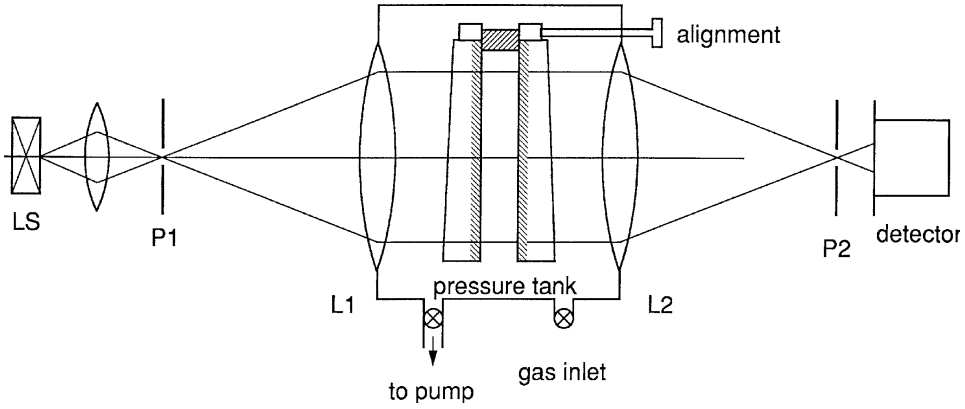

Figure 4.48 Use of a plane FPI for photoelectric recording of the spectrally resolved transmitted intensity $I_T(n \cdot d, \lambda)$ emitted from a point source

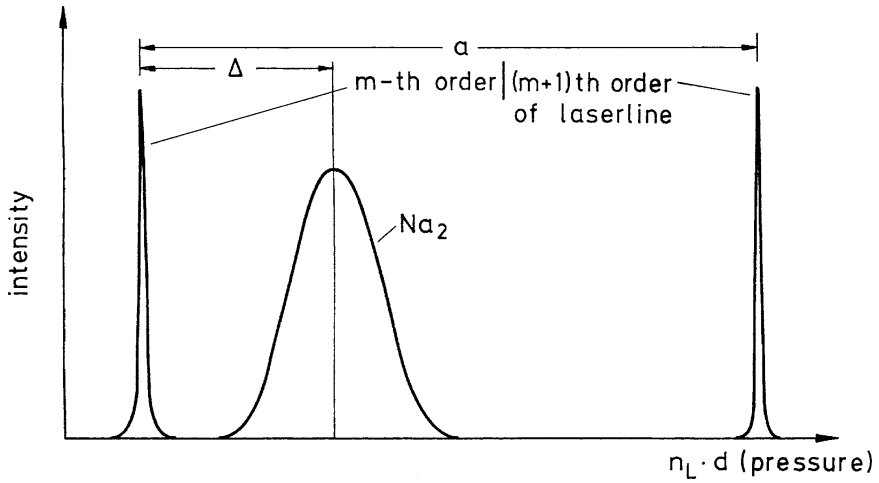

Figure 4.49 Photoelectric recording of a Doppler-broadened laser-excited fluorescence line of Na_2 molecules in a vapor cell and the Doppler-free scattered laser line. The pressure scan $\Delta p = a$ corresponds to one free spectral range of the FPI

FPI, an unambigious determination of λ is possible. For illustration, Fig. 4.49 shows a Doppler-broadened fluorescence line of Na_2 molecules excited by a single-mode argon laser at $\lambda = 488\,\text{nm}$, together with the narrow line profile of the scattered laser light. The pressure change $\Delta p \mathrel{\hat{=}} 2d\,\Delta n_L = a$ corresponds to one free spectral range of the FPI, i.e., $2d\,\Delta n_L = \lambda$.

For Doppler-free resolution of fluorescence lines (Vol. 2, Chap. 4), the laser-induced fluorescence of molecules in a collimated molecular beam can be imaged through a FPI onto the entrance slit of the monochromator (Fig. 4.50). In this case, the crossing point of laser and molecular beam, indeed, represents nearly a point source.

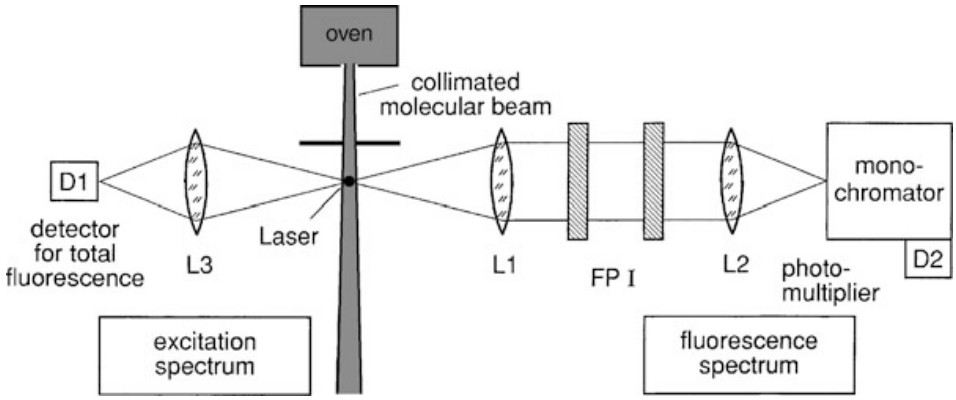

Figure 4.50 Experimental arrangement for photoelectric recording of high-resolution fluorescence lines excited by a single-mode laser in a collimated molecular beam and observed through FPI plus monochromator

4.2.8 Confocal Fabry–Perot Interferometer

A confocal interferometer, sometimes called incorrectly a *spherical FPI*, consists of two spherical mirrors M_1, M_2 with equal curvatures (radius r) that are opposed at a distance $d = r$ (Fig. 4.51a) [152, 153]. These interferometers have gained great importance in laser physics as high-resolution spectrum analyzers for detecting the mode structure and the linewidth of lasers [154–156], and, in the nearly confocal form, as laser resonators (Sect. 5.2).

Neglecting spherical aberration, all light rays entering the interferometer parallel to its axis would pass through the focal point F and would reach the entrance point P1 again after having passed the confocal FPI four times (Fig. 4.51a). Figure 4.51b illustrates the general case of a ray which enters the confocal FPI at a small inclination θ and passes the successive points P1, A, B, C, P1, shown in Fig. 4.51d in a projection. Angle θ is the skew angle of the entering ray.

Because of spherical aberration, rays with different distances ρ_1 from the axis will not all go through F but will intersect the axis at different positions F′ depending on ρ_1 and θ. Also, each ray will not exactly reach the entrance point P_1 after four passages through the confocal FPI since it is slightly shifted at successive passages. However, it can be shown [152, 155] that for sufficiently small angles θ, all rays intersect at a distance $\rho(\rho_1, \theta)$ from the axis in the vicinity of the two points P and P′ located in the central plane of the confocal FPI (Fig. 4.51b).

The optical path difference Δs between two successive rays passing through P can be calculated from geometrical optics. For $\rho_1 \ll r$ and $\theta \ll 1$, one obtains for the near confocal case $d \approx r$ [155]

$$\Delta s = 4d + \rho_1^2 \rho_2^2 \cos 2\theta / r^3 + \text{higher-order terms} . \tag{4.75}$$

An incident light beam with diameter $D = 2\rho_1$ therefore produces, in the central plane of a confocal FPI, an interference pattern of concentric rings. Analogous of

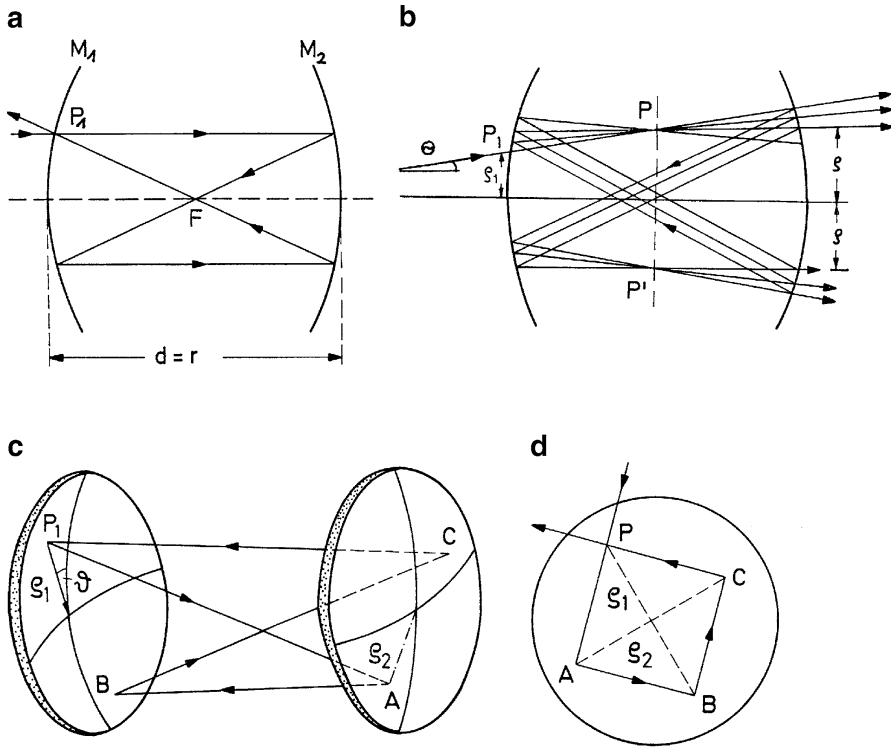

Figure 4.51 Trajectories of rays in a confocal FPI: **a** incident beam parallel to the FPI axis; **b** inclined incident beam; **c** perspective view for illustrating the skew angle; **d** projection of the skewed rays onto the mirror surfaces

the treatment in Sect. 4.2.7, the intensity $I(\rho, \lambda)$ is obtained by adding all amplitudes with their correct phases $\delta = \delta_0 + (2\pi/\lambda)\Delta s$. According to (4.52a), (4.52b) we get

$$I(\rho, \lambda) = \frac{I_0 T^2}{(1-R)^2 + 4R \sin^2[(\pi/\lambda)\Delta s]} , \tag{4.76}$$

where $T = 1 - R - A$ is the transmission of each of the two mirrors. The intensity has maxima for $\delta = 2m\pi$, which is equivalent to

$$4d + \rho^4/r^3 = m\lambda , \tag{4.77}$$

when we neglect the higher-order terms in (4.75) and set $\theta = 0$ and $\rho^2 = \rho_1 \rho_2$.

The free spectral range $\delta\nu$, i.e., the frequency separation between successive interference maxima, is for the near-confocal FPI with $\rho \ll d$

$$\delta\nu = \frac{c}{4d + \rho^4/r^3} , \tag{4.78}$$

which is *different* from the expression $\delta\nu = c/2d$ for the plane FPI.

The radius ρ_m of the mth-order interference ring is obtained from (4.77),

$$\rho_m = [(m\lambda - 4d)r^3]^{1/4} , \tag{4.79}$$

which reveals that ρ_m depends critically on the separation d of the spherical mirrors. Changing d by a small amount ϵ from $d = r$ to $d = r + \epsilon$ changes the path difference to

$$\Delta s = 4(r + \epsilon) + \rho^4/(r + \epsilon)^3 \sim 4(r + \epsilon) + \rho^4/r^3 \,. \tag{4.80}$$

For a given wavelength λ, the value of ϵ can be chosen such that $4(r + \epsilon) = m_0\lambda$. In this case, the radius of the central ring becomes zero. We can number the outer rings by the integer p and obtain with $m = m_0 + p$ for the radius of the pth ring the expression

$$\rho_p = (p\lambda r^3)^{1/4} \,. \tag{4.81}$$

The radial dispersion deduced from (4.79),

$$\frac{\mathrm{d}\rho}{\mathrm{d}\lambda} = \frac{mr^3/4}{[(m\lambda - 4d)r^3]^{3/4}} \,, \tag{4.82}$$

becomes infinite for $m\lambda = 4d$, which occurs according to (4.79) at the center with $\rho = 0$.

This large dispersion can be used for high-resolution spectroscopy of narrow line profiles with a scanning confocal FPI and photoelectric recording (Fig. 4.52).

If the central plane of the near-confocal FPI is imaged by a lens onto a circular aperture with sufficiently small radius $b < (\lambda r^3)^{1/4}$ only the central interference order is transmitted to the detector while all other orders are stopped. Because of the large radial dispersion for small ρ one obtains a high spectral resolving power. With this arrangement not only spectral line profiles but also the instrumental bandwidth can be measured, when an incident monochromatic wave (from a stabilized single-mode laser) is used. The mirror separation $d = r + \epsilon$ is varied by the small amount ϵ and the power

$$P(\lambda, b, \epsilon) = 2\pi \int_{\rho=0}^{b} \rho I(\rho, \lambda, \epsilon)\mathrm{d}\rho \,, \tag{4.83}$$

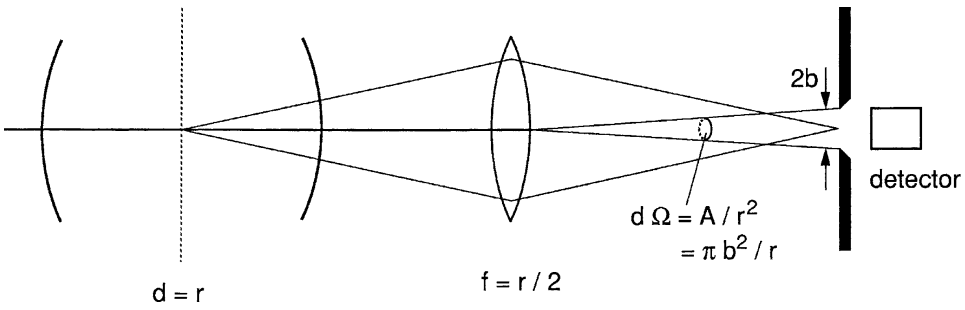

Figure 4.52 Photoelectric recording of the spectral light power transmitted of a scanning confocal FPI

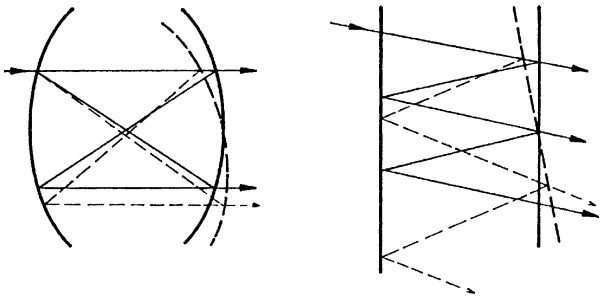

Figure 4.53 Illustration of the larger sensitivity against misalignment for the plane FPI compared with the spherical FPI

transmitted through the aperture is measured as a function of ϵ at fixed values of λ and b.

The integrand $I(\rho, \lambda, \epsilon)$ can be obtained from (4.76), where the phase difference $\delta(\epsilon) = 2\pi\Delta s/\lambda$ is deduced from (4.80).

The optimum choice for the radius b of the aperture is based on a compromise between spectral resolution and transmitted intensity. When the interferometer has the finesse F^*, the spectral halfwidth of the transmission peak is $\delta\nu/F^*$, see (4.55b), and the maximum spectral resolving power becomes $F^*\Delta s/\lambda$ (4.60). For the radius $b = (r^3\lambda/F^*)^{1/4}$ of the aperture, which is just $(F^*)^{1/4}$ times the radius ρ_1 of a fringe with $p = 1$ in (4.81), the spectral resolving power is reduced to about 70 % of its maximum value. This can be verified by inserting this value of b into (4.83) and calculating the halfwidth of the transmission peak $P(\lambda_1, F^*, \epsilon)$.

The total finesse of the confocal FPI is, in general, higher than that of a plane FPI for the following reasons:

- The alignment of spherical mirrors is far less critical than that of plane mirrors, because tilting of the spherical mirrors does not change (to a first approximation) the optical path length $4r$ through the confocal FPI, which remains approximately the same for all incident rays (Fig. 4.53). For the plane FPI, however, the path length increases for rays below the interferometer axis, but decreases for rays above the axis.
- Spherical mirrors can be polished to a higher precision than plane mirrors. This means that the deviations from an ideal sphere are less for spherical mirrors than those from an ideal plane for plane mirrors. Furthermore, such deviations do not wash out the interference structure but cause only a distortion of the ring system because a change of d allows the same path difference Δs for another value of ρ according to (4.75).

The total finesse of a confocal FPI is therefore mainly determined by the reflectivity R of the mirrors. For $R = 0.99$, a finesse $F^* = \pi\sqrt{R/(1-R)} \approx 300$ can be achieved, which is much higher than that obtainable with a plane FPI, where other factors decrease F^*. With the mirror separation $r = d = 3\,\text{cm}$, the free spectral range is $\delta = 2.5\,\text{GHz}$ and the spectral resolution is $\Delta\nu = 7.5\,\text{MHz}$ at the finesse $F^* = 300$. This is sufficient to measure the natural linewidth of many optical transitions. With modern high-reflection coatings, values of $R = 0.9995$ can be obtained and confocal FPI with a finesse $F^* \geq 10^4$ have been realized [157].

From Fig. 4.52 we see that the solid angle accepted by the detector behind the aperture with radius b is $\Omega = \pi b^2/r^2$. The light power transmitted to the detector is proportional to the product of the solid angle Ω and area A in the central plane, which is imaged by the lens onto the aperture (often called the *étendue* U). With the aperture radius $b = (r^3\lambda/F^*)^{1/4}$ (see above) the étendue becomes

$$U = A\Omega = \pi^2 b^4/r^2 = \pi^2 r\lambda/F^* \,. \tag{4.84}$$

For a given finesse F^*, the étendue of the confocal FPI increases with the mirror separation $d = r$. The spectral resolving power

$$\frac{\nu}{\Delta\nu} = 4F^*\frac{r}{\lambda} \,, \tag{4.85}$$

of the confocal FPI is proportional to the product of finesse F^* and the ratio of mirror separation $r = d$ to the wavelength λ. With a given étendue $U = \pi^2 r\lambda/F^*$, we can insert $r = UF^*/(\pi^2\lambda)$ into (4.84) and obtain for the spectral resolving power

$$\frac{\nu}{\Delta\nu} = \left(\frac{2F^*}{\pi\lambda}\right)^2 U \,, \quad \text{(confocal FPI)} \,. \tag{4.86}$$

Let us compare this with the case of a plane FPI with the plate diameter D and the separation d, which is illuminated with nearly parallel light (Fig. 4.48). According to (4.66), the path difference between a ray parallel to the axis and a ray with a small inclination β is, given by $\Delta s = 2nd(1-\cos\beta) \approx nd\beta^2$.

To achieve a finesse F^* with photoelectric recording, this variation of the path length for the different rays through the interferometer should not exceed λ/F^*, which restricts the solid angle $\Omega = \beta^2$ acceptable by the detector to $\Omega \le \lambda/(d \cdot F^*)$. The étendue is therefore

$$U = A\Omega = \pi\frac{D^2}{4}\frac{\lambda}{d \cdot F^*} \,. \tag{4.87}$$

Inserting the value of d given by this equation into the spectral resolving power $\nu/\Delta\nu = 2dF^*/\lambda$, we obtain

$$\frac{\nu}{\Delta\nu} = \frac{\pi D^2}{2U} \,, \quad \text{(plane FPI)} \,. \tag{4.88}$$

While the spectral resolving power is proportional to U for the confocal FPI, it is *inversely proportional to U for the plane FPI*. This is because the étendue increases with the mirror separation d for the confocal FPI but decreases proportional to $1/d$ for the plane FPI. For a mirror radius $r > D^2/4d$, the étendue of the confocal FPI is larger than that of a plane FPI with equal spectral resolution. This means that the transmitted power is larger for the confocal FPI for $r > D^2/4d$.

Example 4.15
A confocal FPI with $r = d = 5\,\text{cm}$ has for $\lambda = 500\,\text{nm}$ the étendue $U = (2.47 \times 10^{-3}/F^*)\,\text{cm}^2/\text{sr}$. This is the same étendue as that of a plane FPI with $d = 5\,\text{cm}$ and $D = 10\,\text{cm}$. However, the diameter of the spherical mirrors can be much smaller (less than 5 mm). With a finesse $F^* = 100$, the étendue is $U = 2.5 \times 10^{-5}$ [$\text{cm}^2\,\text{sr}$] and the spectral resolving power is $\nu/\Delta\nu = 4 \times 10^7$. With this étendue the resolving power of the plane FPI is 6×10^6, provided the whole plane mirror surface has a surface quality to allow a surface finesse of $F^* \geq 100$. In practice, this is difficult to achieve for a flat plane with $D = 10\,\text{cm}$ diameter, while for the small spherical mirrors even $F^* > 10^4$ is feasible.

This example shows that for a given light-gathering power, the confocal FPI can have a much higher spectral resolving power than the plane FPI.

More detailed information on the history, theory, practice, and application of plane and spherical Fabry–Perot interferometers may be found in [158–160].

4.2.9 *Multilayer Dielectric Coatings*

The constructive interference found for the reflection of light from plane-parallel interfaces between two regions with different refractive indices can be utilized to produce highly reflecting, essentially absorption-free mirrors. The improved technology of such dielectric mirrors has greatly supported the development of visible and ultraviolet laser systems.

The reflectivity R of a plane interface between two regions with complex refractive indices $n_1 = n_1' - i\kappa_1$ and $n_2 = n_2' - i\kappa_2$ can be calculated from Fresnel's formulas [129]. It depends on the angle of incidence α and on the direction of polarization. For the polarization component with the electric field vector $\boldsymbol{E}$ parallel to the plane of incidence (defined by the incident and the reflected beam), the reflectivity is

$$R_\text{p} = \left(\frac{n_2 \cos\alpha - n_1 \cos\beta}{n_2 \cos\alpha + n_1 \cos\beta}\right)^2 = \left[\frac{\tan(\alpha - \beta)}{\tan(\alpha + \beta)}\right]^2 \tag{4.89a}$$

where β is the refraction angle ($\sin\beta = (n_1/n_2)\sin\alpha$). For the vertical component ($\boldsymbol{E}$ perpendicular to the plane of incidence), one obtains:

$$R_\text{s} = \left(\frac{n_1 \cos\alpha - n_2 \cos\beta}{n_1 \cos\alpha + n_2 \cos\beta}\right)^2 = \left[\frac{\sin(\alpha - \beta)}{\sin(\alpha + \beta)}\right]^2 . \tag{4.89b}$$

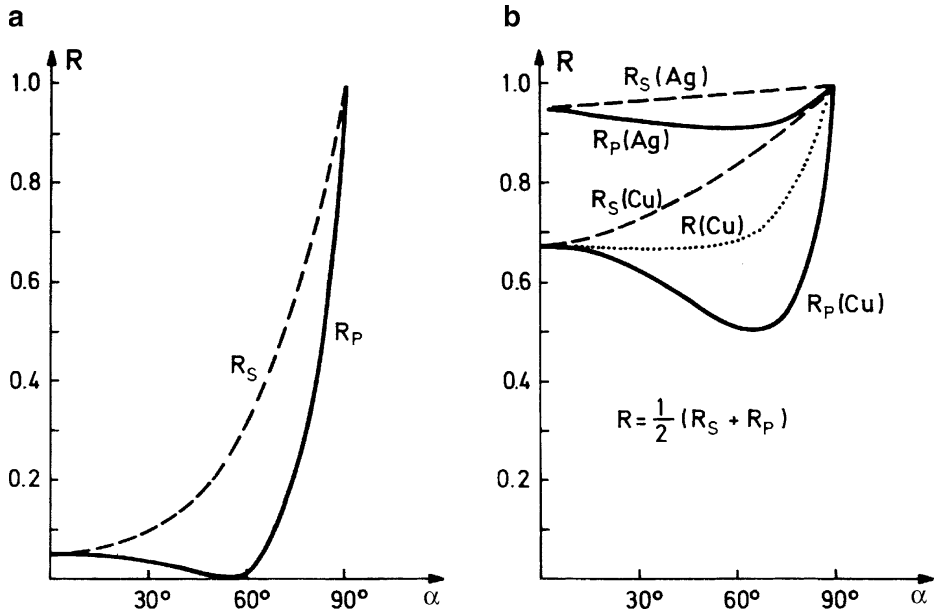

Figure 4.54 Reflectivities R_p and R_s for the two polarization components parallel and perpendicular to the plane of incidence as a function of the angle of incidence α: **a** air–glass boundary ($n_1 = 1, n_2 = 1,5$); **b** air-metal boundary for Cu($n' = 0.76, \kappa = 3.32$) and Ag($n' = 0.055, \kappa = 3.32$)

The reflectivities R_p and R_s are illustrated in Fig. 4.54 for three different materials for incident light polarized parallel (R_p) and perpendicular (R_s) to the plane of incidence.

For vertical incidence ($\alpha = 0, \beta = 0$), one obtains from Fresnel's formulas for both polarizations

$$R|_{\alpha=0} = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 . \tag{4.89c}$$

Since this case represents the most common situation for laser mirrors, we shall restrict the following discussion to vertical incidence.

To achieve maximum reflectivities, the numerator $(n_1 - n_2)^2$ should be maximized and the denominator minimized. Since n_1 is always larger than one, this implies that n_2 should be as large as possible. Unfortunately, the dispersion relations (3.24a, 3.24b) imply that a large value of n also causes large absorption. For instance, highly polished metal surfaces have a maximum reflectivity of $R = 0.95$ in the visible spectral range. The residual 5 % of the incident intensity are absorbed and therefore lost.

The situation can be improved by selecting reflecting materials with low absorption (which then necessarily also have low reflectivity), but using many layers with alternating high and low refractive index n. Choosing the proper optical thickness nd of each layer allows constructive interference between the different reflected amplitudes to be achieved. Reflectivities of up to $R = 0.9999$ have been reached [161–164].

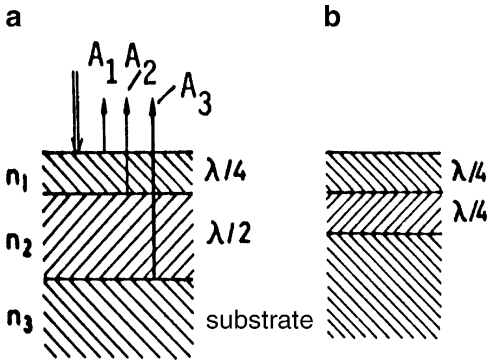

Figure 4.55 Maximum reflection of light with wavelength λ by a two-layer dielectric coating: **a** $n_1 > n_2 > n_3$; **b** $n_1 > n_2 < n_3$

Figure 4.55 illustrates such constructive interference for the example of a two-layer coating. The layers with refractive indices n_1, n_2 and thicknesses d_1, d_2 are evaporated onto an optically smooth substrate with the refractive index n_3. The phase differences between all reflected components have to be $\delta_m = 2m\pi$ ($m = 1, 2, 3, \ldots$) for constructive interference. Taking into account the phase shift $\delta = \pi$ at reflection from an interface with a larger refractive index than that of the foregoing layer, we obtain the conditions

$$n_1 d_1 = \lambda/4 \quad \text{and} \quad n_2 d_2 = \lambda/2 \quad \text{for} \quad n_1 > n_2 > n_3 \,, \tag{4.90a}$$

and

$$n_1 d_1 = n_2 d_2 = \lambda/4 \quad \text{for} \quad n_1 > n_2, n_3 > n_2 \,. \tag{4.90b}$$

The reflected amplitudes can be calculated from Fresnel's formulas. The total reflected intensity is obtained by summation over all reflected amplitudes taking into account the correct phase. The refractive indices are now selected such that $\sum A_i$ becomes a maximum. The calculation is still feasible for our example of a two-layer coating and yields for the three reflected amplitudes (double reflections are neglected)

$$A_1 = \sqrt{R_1} A_0 \,; \quad A_2 = \sqrt{R_2}(1 - \sqrt{R_1}) A_0 \,,$$
$$A_3 = \sqrt{R_3}(1 - \sqrt{R_2})(1 - \sqrt{R_1}) A_0 \,,$$

where the reflectivities R_i are given by (4.89a)–(4.89c).

Example 4.16
$|n_1| = 1.6$, $|n_2| = 1.2$, $|n_3| = 1.45$; $A_1 = 0.231 A_0$, $A_2 = 0.143 A_0$, $A_3 = 0.094 A_0$. $A_R = \sum A_i = 0.468 A_0 \rightarrow I_R = 0.22 I_0 \rightarrow R = 0.22$, provided the path differences have been choosen correctly.

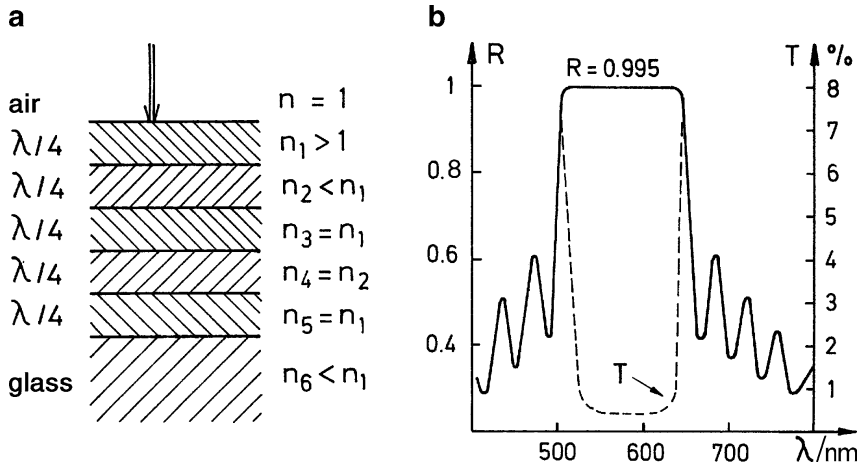

Figure 4.56 The dielectric multilayer mirror: **a** Composition of multilayers; **b** Reflectivity of a high-reflectance multilayer mirror with 17 layers as a function of the incident wavelength λ

This example illustrates that for materials with low absorption, many layers are necessary to achieve a high reflectivity. Figure 4.56a depicts schematically the composition of a dielectric multilayer mirror. The calculation and optimization of multilayer coatings with up to 20 layers becomes very tedious and time consuming, and is therefore performed using computer programs [162, 164]. Figure 4.56b illustrates the reflectivity $R(\lambda)$ of a high-reflectance mirror with 17 layers.

By proper selection of different layers with slightly different optical path lengths, one can achieve a high reflectivity over an extended spectral range. Currently, "broad-band" reflectors are available with reflectivity of $R \geq 0.99$ within the spectral range ($\lambda_0 \pm 0.2\lambda_0$), while the absorption losses are less than 0.2 % [161, 163]. At such low absorption losses, the scattering of light from imperfect mirror surfaces may become the major loss contribution. When total losses of less than 0.5 % are demanded, the mirror substrate must be of high optical quality (better than $\lambda/20$), the dielectric layers have to be evaporated very uniformly, and the mirror surface must be clean and free of dust or dirty films [164]. The best mirrors are produced by ion implantation techniques. Such dielectric mirrors with alternating $\lambda/4$-layers of materials with high and low refractive indices are often called "Bragg mirrors" because they work in a similar way to the Bragg reflection of X-rays at perfect crystal planes. With very pure materials of extremely low absorption, they reach reflectivities of $R > 0.99999$ [165]. The reflectivity $R(\lambda)$ of a Bragg mirror for vertical incidence around $\lambda = 1000\,\text{nm}$ is shown in Fig. 4.57.

Instead of maximizing the reflectivity of a dielectric multilayer coating through *constructive* interference, it is, of course, also possible to minimize it by destructive interference. Such *antireflection coatings* are commonly used to minimize unwanted reflections from the many surfaces of multiple-lens camera objectives, which would otherwise produce an annoying background illumination of the photomaterial. In laser spectroscopy such coatings are important for minimizing reflection losses of optical components inside the laser resonator and for avoiding

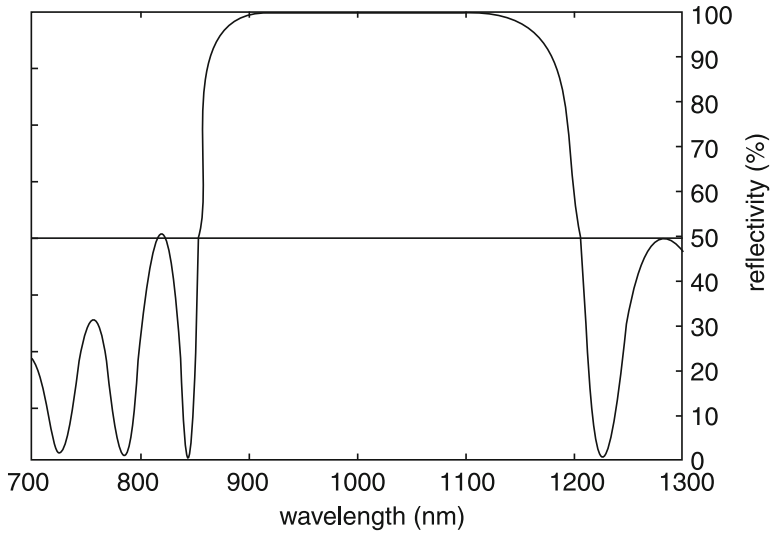

Figure 4.57 Bragg mirror with eight alternating layers of TiO_2 and SiO_2

reflections from the back surface of output mirrors, which would introduce undesirable couplings, thereby causing frequency instabilities of single-mode lasers.

Using a single layer (Fig. 4.58a), the reflectivity reaches a minimum only for a selected wavelength λ (Fig. 4.59). We obtain $I_R = 0$ for $\delta = (2m + 1)\pi$, if the two amplitudes A_1 and A_2 reflected by the interfaces (n_1, n_2) and (n_2, n_3) are equal. For vertical incidence this gives the condition

$$R_1 = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 = R_2 = \left(\frac{n_2 - n_3}{n_2 + n_3}\right)^2 , \tag{4.91}$$

which can be reduced to

$$n_2 = \sqrt{n_1 n_3} . \tag{4.92}$$

For a single layer on a glass substrate the values are $n_1 = 1$ and $n_3 = 1.5$. According to (4.92), n_2 should be $n_2 = \sqrt{1.5} = 1.23$. Durable coatings with such low refractive indices are not available. One often uses MgF_2 with $n_2 = 1.38$, giving a reduction of reflection from 4 % to 1.2 % (Fig. 4.59).

With multilayer antireflection coatings the reflectivity can be decreased below 0.2 % for an extended spectral range [164]. For instance, with three $\lambda/4$ layers

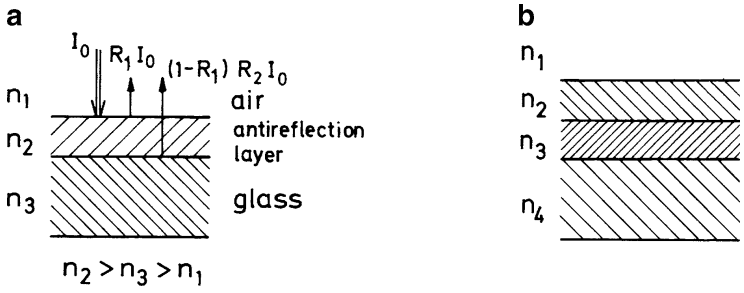

Figure 4.58 Antireflection coating: **a** single layer; **b** multilayer coating

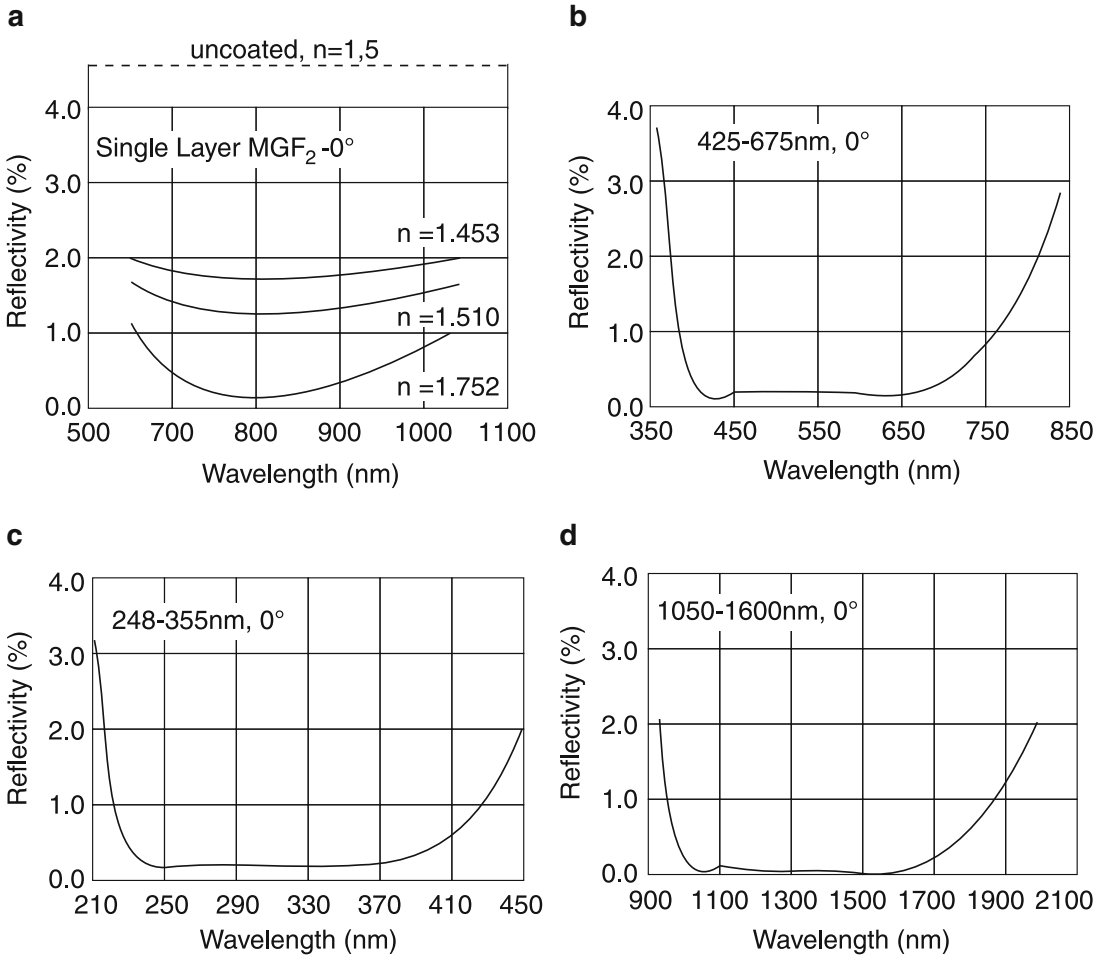

Figure 4.59 Antireflection coatings. **a** Single layer MgF_2 on substrates with different refractive index n; **b–d** broadband multilayer AR-coatings, optimized for different spectral ranges

(MgF_2, SiO, and CeF_3) the reflection drops to below 1 % for the whole range between 420 nm and 840 nm [161, 166, 167].

4.2.10 Interference Filters

Interference filters are used for selective transmission in a narrow spectral range. Incident radiation of wavelengths outside this transmission range is either reflected or absorbed. One distinguishes between line filters and bandpass filters.

A line filter is essentially a Fabry–Perot etalon with a very small optical path nd between the two reflecting surfaces. The technical realization uses two highly reflecting coatings (either silver coatings or dielectric multilayer coatings) that are separated by a nonabsorbing layer with a low refractive index (Fig. 4.60). For instance, for $nd = 0.5\,\mu\text{m}$ the transmission maxima for vertical incidence are obtained from (4.62a) at $\lambda_1 = 1\,\mu\text{m}$, $\lambda_2 = 0.5\,\mu\text{m}$, $\lambda_3 = 0.33\,\mu\text{m}$, etc. In the

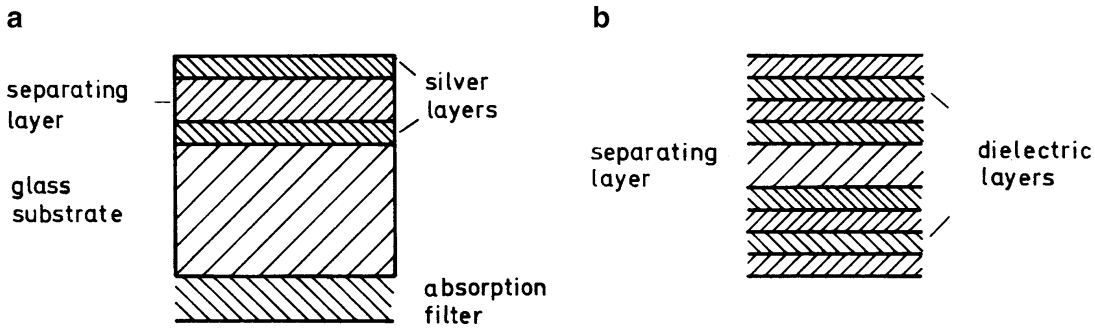

Figure 4.60 Interference filters of the Fabry–Perot type: **a** with two single layers of silver; **b** with dielectric multilayer coatings

visible range this filter has therefore only one transmission peak at $\lambda = 500\,\text{nm}$, with a halfwidth that depends on the finesse $F^* = \pi\sqrt{R}/(1-R)$ (Fig. 4.38).

The interference filter is characterized by the following quantities:

- The wavelength λ_m at peak transmission;
- The maximum transmission;
- The contrast factor, which gives the ratio of maximum to minimum transmission;
- The bandwidth $\Delta\nu = \nu_1 - \nu_2$ with $T(\nu_1) = T(\nu_2) = \frac{1}{2}T_{\max}$.

The maximum transmission according to (4.61a)–(4.61d) is $T_{\max} = T^2/(1-R)^2$. Using thin silver or aluminum coatings with $R = 0.8$, $T = 0.1$, and $A = 0.1$, the transmission of the filter is only $T_{\max} = 0.25$ and the finesse $F^* = 15$. For our example this means a halfwidth of $660\,\text{cm}^{-1}$ at a free spectral range of $10^4\,\text{cm}^{-1}$. At $\lambda = 500\,\text{nm}$ this corresponds to a free spectral range of 250 nm and a halfwidth of about 16 nm. For many applications in laser spectroscopy, the low peak transmission of interference filters with absorbing metal coatings is not tolerable. One has to use absorption-free dielectric multilayer coatings (Fig. 4.60b) with high reflectivity, which allows a large finesse and therefore a smaller bandwidth and a larger peak transmission (Fig. 4.61).

Example 4.17

With $R = 0.95$, $A = 0.01$ and $T = 0.04$, according to (4.61a)–(4.61d) we obtain a peak transmission of 64 %, which increases with $A = 0.005$, $T = 0.045$ to 81 %. The contrast becomes $\gamma = I_T^{\max}/I_T^{\min} = (1+F) = 1 + 4F^{*2}/\pi^2 = 1520$. With a thickness $nd = 5\,\mu\text{m}$ of the separating layer, the free spectral range is $\delta\nu = 3 \times 10^{13}\,\text{Hz} \mathrel{\hat{=}} 25\,\text{nm}$ at $\lambda = 500\,\text{nm}$.

A higher finesse F^* due to larger reflectivities of the reflecting films not only decreases the bandwidth but also increases the contrast factor. With $R = 0.98 \rightarrow F = 4R/(1-R)^2 = 9.8 \times 10^3$, which means that the intensity at the transmission minimum is only about 10^{-4} of the peak transmission.

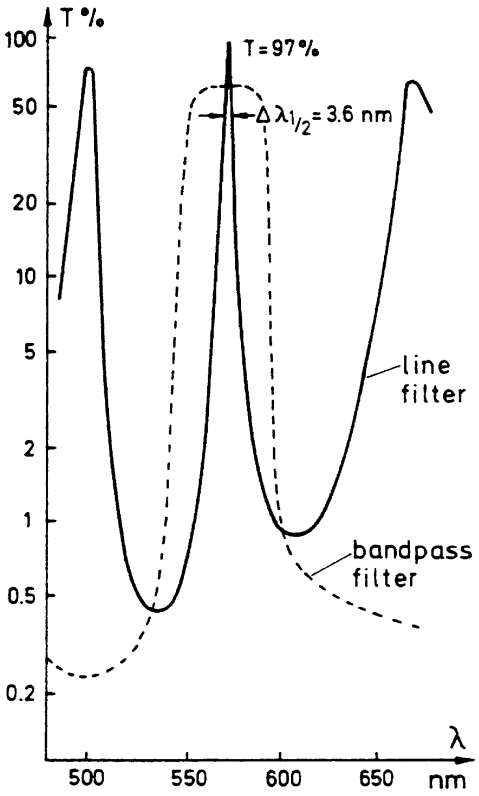

Figure 4.61 Spectral transmission of interference filters. *Solid curve*: line filter. *Dashed curve*: bandpass filter. Note the logarithmic scale

The bandwidth can be further decreased by using two interference filters in series. However, it is preferable to construct a double filter that consists of three highly-reflecting surfaces, separated by two nonabsorbing layers of the same optical thickness. If the thickness of these two layers is made slightly different, a bandpass filter results that has a flat transmission curve but steep slopes to both sides. Commercial interference filters are currently available with a peak transmission of at least 90 % and a bandwidth of less than 2 nm [162, 168]. Special narrow-band filters even reach 0.3 nm, however, with reduced peak transmission.

The wavelength λ_m of the transmission peak can be shifted to lower values by tilting the interference filter, which increases the angle of incidence α, see (4.62a). The tuning range is, however, restricted, because the reflectivity of the multilayer coatings also depends on the angle α and is, in general, optimized for $\alpha = 0$. For divergent incident light, the transmission bandwidth increases with the divergence angle. From (4.62a), we obtain for the wavelength $\lambda(\alpha)$ of a tilted filter

$$\lambda = \frac{2nd}{m}\cos\beta = \lambda_0 \cos\beta \approx \lambda_0\left(1 - \frac{\beta^2}{2}\right) \approx \lambda_0\left(1 - \frac{\alpha^2}{2n^2}\right) . \qquad (4.93)$$

Example 4.18

$\lambda_0 = 1500\,\text{nm}$, $n = 1.5$, $\alpha = 150° \mathrel{\hat{=}} 0.25\,\text{rad} \Rightarrow \lambda(\alpha) = 1389\,\text{nm} \Rightarrow$
$\Delta\lambda = \lambda_0 - \lambda(\alpha) = 111\,\text{nm}$

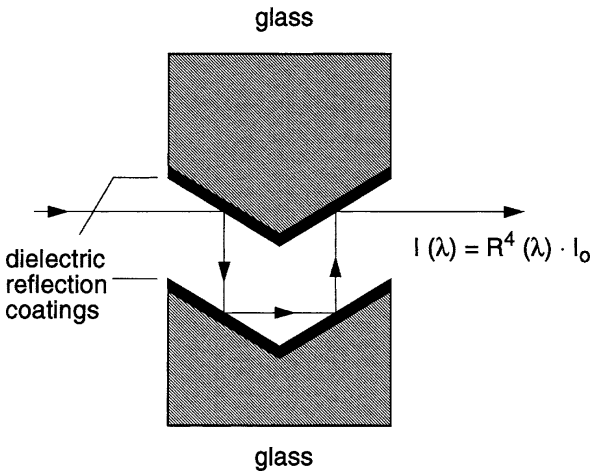

Figure 4.62 Reflection interference filter

In the ultraviolet region, where the absorption of most materials used for interference filters becomes large, the selective *reflectance* of interference filters can be utilized to achieve narrow-band filters with low losses (Fig. 4.62). For more detailed treatment, see [161–168].

In low-level fluorescence spectroscopy or Raman spectroscopy, the scattered light of the intense exciting laser often overlaps the fluorescence lines. Here special interference filters are available which have a narrow minimum transmission at the laser wavelength (line-blocking filter) but a high transmission in the other spectral ranges.

Since temperature drifts cause a change of the spacing d, the wavelength λ_p at peak transmission also shifts with temperature. Typical values are $\mathrm{d}\lambda_\mathrm{p}/\mathrm{d}T = 0.02\,\mathrm{nm/K}$. A temperature change of 10 K therefore shifts the peak transmission by 0.2 nm. This is only relevant for filters with a very narrow transmission bandwidth.

4.2.11 Birefringent Interferometer

The basic principle of the birefringent interferometer or *Lyot filter* [137, 169] is founded on the interference of polarized light that has passed through a birefringent crystal. Assume that a linearly polarized plane wave

$$\boldsymbol{E} = \boldsymbol{A} \cdot \cos(\omega t - kx) \,,$$

with

$$\boldsymbol{A} = \{0, A_y, A_z\} \,, \quad A_y = |A| \sin\alpha \,, \quad A_z = |A| \cos\alpha \,,$$

is incident on the birefringent crystal (Fig. 4.63). The electric vector $\boldsymbol{E}$ makes an angle α with the optical axis, which points into the z-direction. Within the crystal, the wave is split into an ordinary beam with the wave number $k_\mathrm{o} = n_\mathrm{o} k$

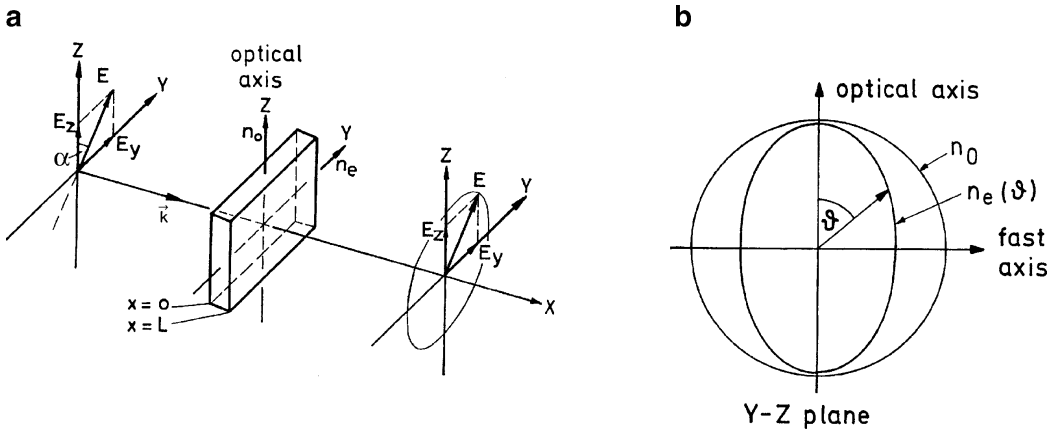

Figure 4.63 Lyot filter: **a** schematic arrangement; **b** index ellipsoid of the birefringent crystal

and the phase velocity $v_o = c/n_o$, and an extraordinary beam with $k_\mathrm{e} = n_\mathrm{e}k$ and $v_\mathrm{e} = c/n_\mathrm{e}$. The partial waves have mutually orthogonal polarization in directions parallel to the z- and y-axis, respectively. Let the crystal with length L be placed between $x = 0$ and $x = L$. Because of the different refractive indices n_o and n_e for the ordinary and the extraordinary beams, the two partial waves at $x = L$

$$E_y(L) = A_y \cos(\omega t - k_\mathrm{e} L) \quad \text{and} \quad E_z(L) = A_z \cos(\omega t - k_0 L) \,,$$

show a phase difference of

$$\Delta\phi = k(n_0 - n_\mathrm{e})L = (2\pi/\lambda)\Delta n L \quad \text{with } \Delta n = n_0 - n_\mathrm{e} \,. \tag{4.94}$$

The superposition of these two waves results, in general, in elliptically polarized light, where the principal axis of the ellipse is turned by an angle $\beta = \phi/2$ against the direction of $\boldsymbol{A}_0$.

For phase differences $\Delta\phi = 2m\pi$, linearly polarized light with $\boldsymbol{E}(L) \parallel \boldsymbol{E}(0)$ is obtained. However, for $\Delta\phi = (2m+1)\pi$ and $\alpha = 45°$, the transmitted wave is also linearly polarized, but now $\boldsymbol{E}(L) \perp \boldsymbol{E}(0)$.

The elementary Lyot filter consists of a birefringent crystal placed between two linear polarizers (Fig. 4.63a). Assume that the two polarizers are both parallel to the electric vector $\boldsymbol{E}(0)$ of the incoming wave. The second polarizer parallel to $\boldsymbol{E}(0)$ transmits only the projection

$$\begin{aligned} E &= E_y \sin\alpha + E_z \cos\alpha \\ &= A[\sin^2\alpha \cos(\omega t - k_\mathrm{e} L) + \cos^2\alpha \cos(\omega t - k_0 L)] \,, \end{aligned}$$

of the amplitudes, which yields with (4.91) the transmitted time averaged intensity

$$\bar{I}_\mathrm{T} = \tfrac{1}{2} c\epsilon_0 \overline{E^2} = \bar{I}_0(\sin^4\alpha + \cos^4\alpha + 2\sin^2\alpha\cos^2\alpha\cos\Delta\phi) \,. \tag{4.95}$$

Using the relations $\cos\phi = 1 - 2\sin^2\frac{1}{2}\phi$, and $2\sin\alpha\cos\alpha = \sin 2\alpha$, this reduces to

$$\bar{I}_\mathrm{T} = I_0[1 - \sin^2\tfrac{1}{2}\Delta\phi\,\sin^2(2\alpha)]\ , \tag{4.96}$$

which gives for $\alpha = 45°$

$$I_\mathrm{T} = I_0\left[1 - \sin^2\frac{\Delta\phi}{2}\right] = I_0\cos^2\frac{\Delta\phi}{2}\ . \tag{4.96a}$$

The transmission of the Lyot filter is therefore a function of the phase retardation, i.e.,

$$\boxed{T(\lambda) = \frac{I_\mathrm{T}}{I_0} = T_0\cos^2\left(\frac{\pi\Delta n L}{\lambda}\right)} \tag{4.97}$$

which depends on the wavelength λ.

Note According to (4.96) the maximum modulation of the transmittance with $T_\mathrm{max} = T_0$ and $T_\mathrm{min} = 0$ is only achieved for $\alpha = 45°$!

Taking into account absorption and reflection losses, the maximum transmission $I_\mathrm{T}/I_0 = T_0 < 1$ becomes less than 100 %. Within a small wavelength interval, the difference $\Delta n = n_0 - n_\mathrm{e}$ can be regarded as constant. Therefore (4.97) gives the wavelength-dependent transmission function, $\cos^2\phi$, typical of a two-beam interferometer (Fig. 4.26). For extended spectral ranges the different dispersion of $n_\mathrm{o}(\lambda)$ and $n_\mathrm{e}(\lambda)$ has to be considered, which causes a wavelength dependence, $\Delta n(\lambda)$.

The free spectral range $\delta\nu$ is obtained from (4.97) as

$$\frac{\Delta n\cdot L}{\lambda_1} - \frac{\Delta n\cdot L}{\lambda_2} = 1\ .$$

With $\nu = c/\lambda$, this becomes

$$\delta\nu = \frac{c}{(n_\mathrm{o} - n_\mathrm{e})L}\ . \tag{4.98}$$

Example 4.19
For a crystal of potassium dihydrogen phosphate (KDP), $n_\mathrm{e} = 1.51$, $n_0 = 1.47 \rightarrow \Delta n = 0.04$ at $\lambda = 600\,\mathrm{nm}$. A crystal with $L = 2\,\mathrm{cm}$ then has a free spectral range $\delta\nu = 3.75\times10^{11}\,\mathrm{Hz} \mathrel{\widehat{=}} \delta\bar{\nu} = 12.5\,\mathrm{cm}^{-1} \rightarrow \Delta\lambda = 0.45\,\mathrm{nm}$ at $\lambda = 600\,\mathrm{nm}$.

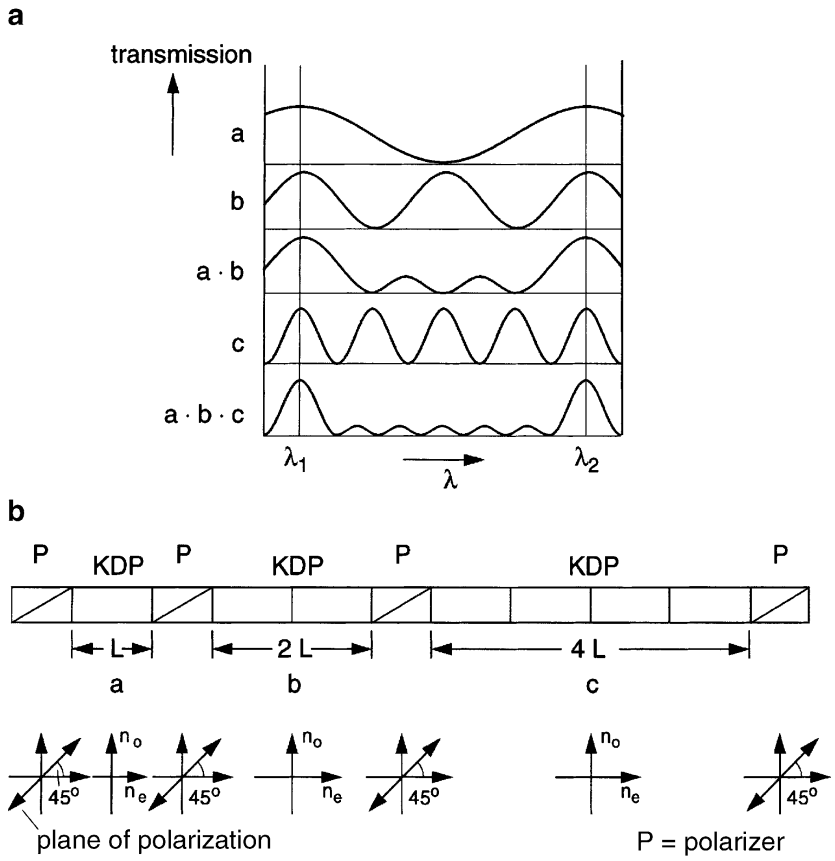

Figure 4.64 **a** Transmitted intensity $I_T(\lambda)$ of a Lyot filter composed of three birefringent crystals with lengths L, $2L$, and $4L$ between polarizers. **b** Arrangement of the crystals and the state of polarization of the transmitted wave

If N elementary Lyot filters with different lengths L_m are placed in series, the total transmission T is the product of the different transmissions T_m, i.e.,

$$T(\lambda) = \prod_{m=1}^{N} T_{0m} \cos^2\left(\frac{\pi \Delta n L_m}{\lambda}\right) . \tag{4.99}$$

Figure 4.64 illustrates a possible experimental arrangement and the corresponding transmission for a Lyot filter composed of three components with the lengths $L_1 = L$, $L_2 = 2L$, and $L_3 = 4L$. The free spectral range $\delta\nu$ of this filter equals that of the shortest component; the halfwidth $\Delta\nu$ of the transmission peaks is, however, mainly determined by the longest component. When we define, analogous to the Fabry–Perot interferometer, the finesse F^* of the Lyot filter as the ratio of the free spectral range $\delta\nu$ to the halfwidth $\Delta\nu$, we obtain, for a composite Lyot filter with N elements of lengths $L_m = 2^{m-1}L_1$, a finesse that is approximately $F^* = 2^N$.

The wavelength of the transmission peak can be tuned by changing the difference $\Delta n = n_o - n_e$. This can be realized in two different ways:

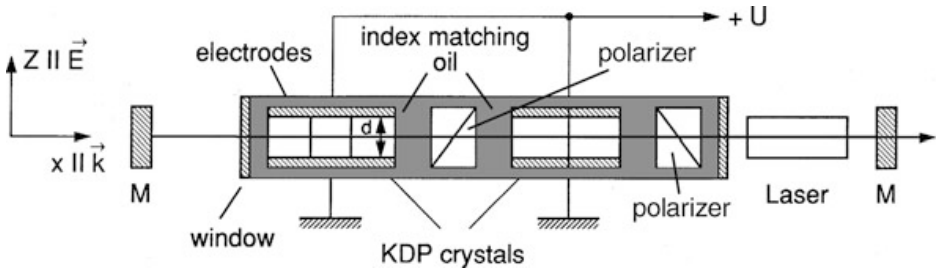

Figure 4.65 Electro-optic tuning of a Lyot filter [170]

- By changing the angle θ between the optical axis and the wave vector $\boldsymbol{k}$, which alters the index n_e. This can be illustrated with the index ellipsoid (Fig. 4.63b), which gives both refractive indices for a given wavelength as a function of θ. The difference $\Delta n = n_o - n_e$ therefore depends on θ. The two axes of the ellipsoid with minimum n_e ($\theta = 90°$ for a negative birefringent crystal) and maximum n_o ($\theta = 0°$) are often called the *fast* and the *slow* axes. Turning the crystal around the x-axis in Fig. 4.63a, which is perpendicular to the y–z-plane of Fig. 4.63b, results in a continuous change of Δn and a corresponding tuning of the peak transmission wavelength λ (Sect. 5.7.4).
- By using the different dependence of the refractive indices n_o and n_e on an applied electric field [171]. This "induced birefringence" depends on the orientation of the crystal axis in the electric field. A common arrangement employs a potassium dihydrogen phosphate (KDP) crystal with an orientation where the electric field is parallel to the optical axis (z-axis) and the wave vector $\boldsymbol{k}$ of the incident wave is perpendicular to the z-direction (transverse electro-optic effect, Fig. 4.65). Two opposite sides of the rectangular crystal with the side length d are coated with gold electrodes and the electric field $E = U/d$ is controlled by the applied voltage.

In the external electric field the uniaxial crystal becomes biaxial. In addition to the natural birefringence of the uniaxial crystal, a field-induced birefringence is generated, which is approximately proportional to the field strength E [172]. The changes of n_o or n_e by the electric field depend on the symmetry of the crystal, the direction of the applied field, and on the magnitude of the electro-optic coefficients. For the KDP crystal only one electro-optic coefficient $d_{36} = -10.7 \times 10^{-12}$ [m/V] (see Sect. 6.1) is effective if the field is applied parallel to the optical axis.

The difference $\Delta n = n_o - n_e$ then becomes

$$\Delta n(E_z) = \Delta n(E = 0) + \tfrac{1}{2} n_1^3 d_{36} E_z \ . \tag{4.100}$$

Maximum transmittance is obtained for

$$\Delta n L = m\lambda \quad (m = 0, 1, 2 \ldots) \ ,$$

which gives the wavelength λ at the maximum transmittance

$$\lambda = (\Delta n(E = 0) + 0.5\, n_1{}^3 d_{36} E_z) L/m \,, \tag{4.101}$$

as a function of the applied field.

While this electro-optic tuning of the Lyot filter allows rapid switching of the peak transmission, for many applications, where a high tuning speed is not demanded, mechanical tuning is more convenient and easier to realize.

4.2.12 Tunable Interferometers

For many applications in laser spectroscopy it is advantageous to have a high-resolution interferometer that is able to scan, in a given time interval Δt, through a limited spectral range $\Delta\nu$. The scanning speed $\Delta\nu/\Delta t$ depends on the method used for tuning, while the spectral range $\Delta\nu$ is limited by the free spectral range $\delta\nu$ of the instrument. All techniques for tuning the wavelength $\lambda_m = 2nd/m$ at the transmission peak of an interferometer are based on a continuous change of the optical path difference between successive interfering beams [171, 173]. This can be achieved in different ways:

a) Change the refractive index n by altering the pressure between the reflecting plates of a FPI (pressure-scanned FPI);
b) Change the distance d between the plates with piezoelectric or magnetostrictive elements;
c) Tilt the solid etalons with a given thickness d against the direction of the incoming plane wave;
d) Change the optical path difference $\Delta s = \Delta n L$ in birefringent crystals by electro-optic tuning or by turning the optical axis of the crystal (Lyot filter).

While method (a) is often used for high-resolution fluorescence spectroscopy with slow scan rates or for tuning pulsed dye lasers, method (b) is realized in a scanning confocal FPI (used as an optical spectrum analyzer) for monitoring the mode structure of lasers.

With a commercial spectrum analyzer, the transmitted wavelength λ can be repetitively scanned over more than one free spectral range with a saw-tooth voltage (Fig. 4.66) applied to the piezoelectric distance holder [154, 173]. Scanning rates up to several kilohertz are possible. Although the finesse of such devices may exceed 10^3, the hysteresis of piezoelectric crystals limits the accuracy of absolute wavelength calibration. Here a pressure-tuned FPI may be advantageous. The pressure change has to be sufficiently slow to avoid turbulence and temperature drifts. With a digitally pressure-scanned FPI, where the pressure of the gas in the interferometer chamber is changed by small, discrete steps, repetitive scans are reproduced within about 10^{-3} of the free spectral range [174].

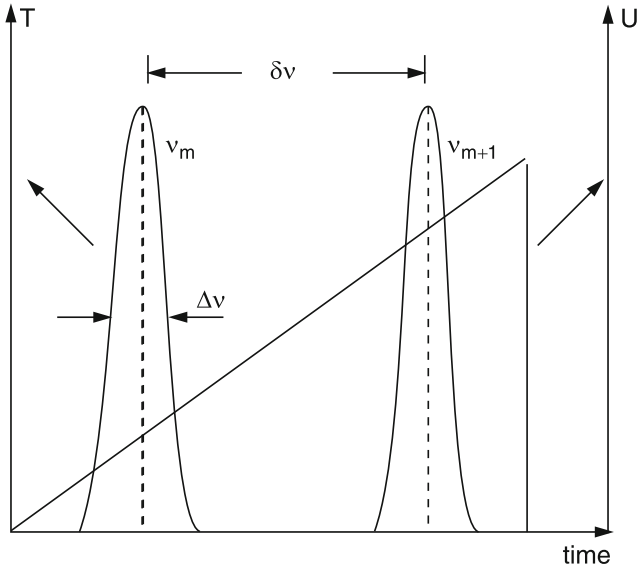

Figure 4.66 Scanning confocal FPI with transmission peaks of a fundamental laser mode and sawtooth voltage at the piezo on one mirror

For fast wavelength tuning of dye lasers, Lyot filters with electro-optic tuning are employed within the laser resonator. A tuning range of a few nanometers can be repetitively scanned with rates up to 10^5 per second [175].

4.3 Comparison Between Spectrometers and Interferometers

When comparing the advantages and disadvantages of different dispersing devices for spectroscopic analysis, the characteristic properties of the instruments discussed in the foregoing sections, such as *spectral resolving power, étendue, spectral transmission*, and *free spectral* range, are important for the optimum choice. Of equal significance is the question of how accurately the wavelengths of spectral lines can be measured. To answer this question, further specifications are necessary, such as the backlash of monochromator drives, imaging errors in spectrographs, and hysteresis in piezo-tuned interferometers. In this section we shall treat these points in a comparison for different devices in order to give the reader an impression of the capabilities and limitations of these instruments.

4.3.1 Spectral Resolving Power

The spectral resolving power discussed for the different instruments in the previous sections can be expressed in a more general way, which applies to all devices with spectral dispersion based on interference effects. Let Δs_m be the maximum path difference between interfering waves in the instrument, e.g., between the rays from the first and the last groove of a grating (Fig. 4.67a) or between the direct beam and a beam reflected m times in a Fabry–Perot interferometer (Fig. 4.67b). Two wave-

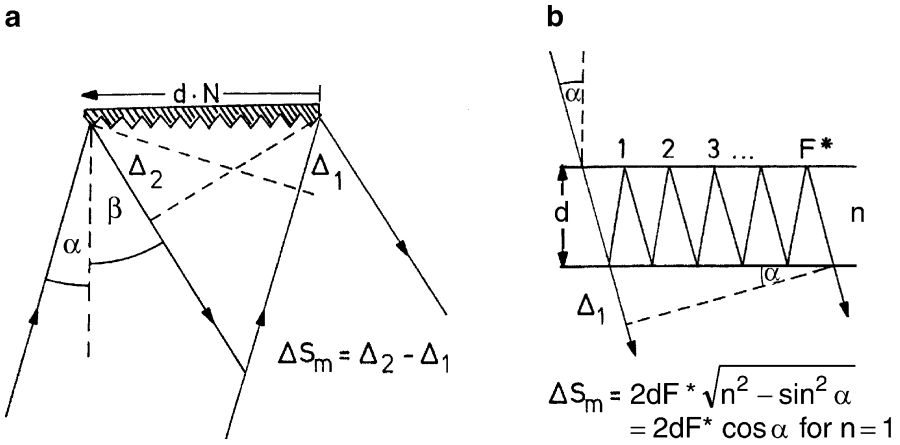

Figure 4.67 Maximum optical path difference and spectral resolving power: **a** in a grating spectrometer; **b** in a Fabry–Perot interferometer

lengths λ_1 and $\lambda_2 = \lambda_1 + \Delta\lambda$ can still be resolved if the number of wavelengths over this maximum path difference

$$\Delta s_\mathrm{m} = 2m\lambda_2 = (2m+1)\lambda_1 \,, \quad m = \text{integer} \,,$$

differs for the two wavelengths by at least one unit. In this case, an interference maximum for λ_1 coincides with the first minimum for λ_2. From the above equation we obtain the theoretical upper limit for the resolving power

$$\boxed{\frac{\lambda}{\Delta\lambda} = \frac{\Delta s_\mathrm{m}}{\lambda} \,,} \tag{4.102}$$

which is equal to the maximum path difference measured in units of the wavelength λ.

With the maximum time difference $\Delta T_\mathrm{m} = \Delta s_\mathrm{m}/c$ for traversing the two paths with the path difference Δs_m, we obtain with $\nu = c/\lambda$ from (4.102) for the minimum resolvable interval $\Delta\nu = -(c/\lambda^2)\Delta\lambda$,

$$\Delta\nu = 1/\Delta T_\mathrm{m} \;\Rightarrow\; \Delta\nu \cdot \Delta T_\mathrm{m} = 1 \,. \tag{4.103}$$

The product of the minimum resolvable frequency interval $\Delta\nu$ *and the maximum difference in transit times through the spectral apparatus is equal to 1.*

Example 4.20

a) **Grating Spectrometer:** The maximum path difference is, according to (4.30) and Fig. 4.67,

$$\Delta s_m = Nd(\sin\alpha - \sin\beta) = mN\lambda \,.$$

The upper limit for the resolving power is therefore, according to (4.102),

$$R = \lambda/\Delta\lambda = mN \quad \begin{matrix}(m : \text{diffraction order},\\ N : \text{number of illuminated grooves})\,.\end{matrix}$$

For $m = 2$ and $N = 10^5$ this gives $R = 2\times10^5$, or $\Delta\lambda = 5\times10^{-6}\lambda$. Because of diffraction, which depends on the size of the grating (Sect. 4.1.3), the realizable resolving power is 2–3 times lower. This means that at $\lambda = 500\,\text{nm}$, two lines with $\Delta\lambda \geq 10^{-2}\,\text{nm}$ can still be resolved.

b) **Michelson Interferometer:** The path difference Δs between the two interfering beams is changed from $\Delta s = 0$ to $\Delta s = \Delta s_\text{m}$. The numbers of interference maxima are counted for the two components λ_1 and λ_2 (Sect. 4.2.4). A distinction between λ_1 and λ_2 is possible if the number $m_1 = \Delta s/\lambda_1$ differs by at least 1 from $m_2 = \Delta s/\lambda_2$; this immediately gives (4.102). With a modern design, maximum path differences Δs up to several meters have been realized for wavelength measurements of stabilized lasers (Sect. 4.5.3). For $\lambda = 500\,\text{nm}$ and $\Delta s = 1\,\text{m}$, we obtain $\lambda/\Delta\lambda = 2\times10^6$, which is one order of magnitude better than for the grating spectrometer.

c) **Fabry–Perot Interferometer:** The path difference is determined by the optical path difference $2nd$ between successive partial beams times the effective number of reflections, which can be expressed by the reflectivity finesse $F^* = \pi\sqrt{R}/(1-R)$. With ideal reflecting planes and perfect alignment, the maximum path difference would be $\Delta s_\text{m} = 2ndF^*$ and the spectral resolving power, according to (4.102), would be

$$\lambda/\Delta\lambda = F^* 2nd/\lambda\,.$$

Because of imperfections of the alignment and deviations from ideal planes, the effective finesse is lower than the reflectivity finesse. With a value of $F^*_\text{eff} = 50$, which can be achieved, we obtain for $nd = 1\,\text{cm}$

$$\lambda/\Delta\lambda = 2\times10^6\,,$$

which is comparable with the Michelson interferometer having $\Delta s_m = 100\,\text{cm}$. However, with a confocal FPI, a finesse of $F^*_\text{eff} = 1000$ can be achieved. With $r = d = 4\,\text{cm}$ we then obtain

$$\lambda/\Delta\lambda = F^* 4d/\lambda \approx 5\times10^8\,,$$

which means that for $\lambda = 500\,\text{nm}$, two lines with $\Delta\lambda = 1\times10^{-6}\,\text{nm}$ ($\Delta\nu = 1\,\text{MHz}$ at $\nu = 5\times10^{14}\,\text{s}^{-1}$) are still resolvable, provided that their linewidth is sufficiently small. With high-reflection mirror coatings a finesse of $F^*_\text{eff} = 10^5$ has been realized. With $r = d = 1\,\text{m}$ this yields $\lambda/\Delta\lambda = 8\times10^{11}$ [160].

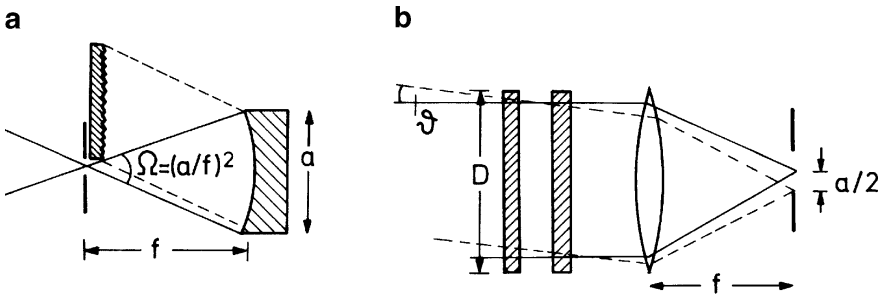

Figure 4.68 Acceptance angle of a spectrometer (**a**); and a Fabry–Perot interferometer (**b**)

4.3.2 *Light-Gathering Power*

The *light-gathering power*, or *étendue*, has been defined in Sect. 4.1.1 as the product $U = A\Omega$ of entrance area A and solid angle of acceptance Ω of the spectral apparatus. For most spectroscopic applications it is desirable to have an étendue U as large as possible to gain intensity. An equally important goal is to reach a maximum resolving power R. However, the two quantitites U and R are not independent of each other but are related, as can be seen from the following examples.

Example 4.21

a) **Spectrometer:** The area of the entrance slit with width b and height h is $A = b \cdot h$. The acceptance angle $\Omega = (a/f)^2$ is determined by the focal length f of the collimating lens or mirror and the diameter a of the limiting aperture in the spectrometer (Fig. 4.68a). We can write the étendue,

$$U = bha^2/f^2 \,,$$

as the product of the area $A = bh$ and the solid angle $\Omega = (a/f)^2$.
Using typical figures for a medium-sized spectrometer ($b = 10\,\mu\text{m}$, $h = 0.5\,\text{cm}$, $a = 10\,\text{cm}$, $f = 100\,\text{cm}$) we obtain $\Omega = 0.01$, $A = 5 \times 10^{-4}\,\text{cm}^2 \rightarrow U = 5 \times 10^{-6}\,\text{cm}^2\,\text{sr}$. With the resolving power $R = mN$, the product

$$RU = mNA\Omega \approx mN\frac{bha^2}{f^2} \,, \tag{4.104a}$$

increases with the diffraction order m, the size a of the grating, the number of illuminated grooves N, and the slit area bh (as long as imaging errors can be neglected). For $m = 1$, $N = 10^5$, and the above figures for h, b, a, and f, we obtain $RU = 0.5\,\text{cm}^2\,\text{sr}$.

b) **Interferometer:** For the Michelson and Fabry–Perot interferometers, the allowable acceptance angle for photoelectric recording is limited by the aperture in front of the detector, which selects the central circular fringe. From Figs. 4.52 and 4.68b we see that the fringe images at the center and at the edge of the limiting aperture with diameter a are produced by incoming beams that are inclined by an angle ϑ against each other. With $a/2 = f\vartheta$, the solid angle accepted by the FPI is $\Omega = a^2/(4f^2)$. For a plate diameter D the étendue is then $U = \pi(D^2/4)\Omega$. According to (4.88) the spectral resolving power $R = \nu/\Delta\nu$ of a plane FPI is correlated with the étendue U by $R = \pi D^2(2U)^{-1}$. The product

$$RU = \pi D^2/2 \,, \tag{4.104b}$$

is, for a plane FPI, therefore solely determined by the plate diameter. For $D = 5\,\text{cm}$, RU is about $40\,\text{cm}^2\,\text{sr}$, and therefore two orders of magnitude larger than for a grating spectrometer.

In Sect. 4.2.12 we saw that for a given resolving power the spherical FPI has a larger étendue for mirror separations $r > D^2/4d$. For Example 4.21 with $D = 5\,\text{cm}$, $d = 1\,\text{cm}$, the confocal FPI therefore gives the largest product RU of all interferometers for $r > 6\,\text{cm}$. Because of the higher total finesse, however, the confocal FPI may be superior to all other instruments even for smaller mirror separations.

In summary, we can say that at comparable resolving power interferometers have a larger lightgathering power than spectrometers.

4.4 Accurate Wavelength Measurements

One of the major tasks for spectroscopists is the measurement of wavelengths of spectral lines. This allows the determination of molecular energy levels and of molecular structure. The attainable accuracy of wavelength measurements depends not only on the spectral resolution of the measuring device but also on the achievable signal-to-noise ratio and on the reproducibility of measured absolute wavelength values.

With the ultrahigh resolution, which can, in principle, be achieved with single-mode tunable lasers (Vol. 2, Chaps. 1–5), the accuracy of absolute wavelength measurements attainable with conventional techniques may not be satisfactory. New methods have been developed that are mainly based on interferometric measurements of laser wavelengths. For applications in molecular spectroscopy, the laser can be stabilized on the center of a molecular transition. Measuring the wavelength of such a stabilized laser yields simultaneously the wavelength of the molecular

transition with a comparable accuracy. We shall briefly discuss some of these devices, often called *wavemeters*, that measure the unknown laser wavelength by comparison with a reference wavelength λ_R of a stabilized reference laser. Most proposals use for reference a HeNe laser, stabilized on a hyperfine component of a molecular iodine line, which has been measured by direct comparison with the primary wavelength standard to an accuracy of better than 10^{-10} [176].

Another method measures the absolute frequency ν_L of a stabilized laser and deduces the wavelength λ_L from the relation $\lambda_L = c/\nu_L$ using the *best* average of experimental values for the speed of light [177–179], which has been chosen to *define* the meter and thus the wavelength λ by the definition: 1 m *is the distance traveled by light in vacuum during the time* $\Delta t = 1/299{,}792{,}458\,\mathrm{s}^{-1}$. *This defines the speed of light* as

$$c = 299{,}792{,}458\,\mathrm{m/s}\ . \tag{4.105}$$

Such a scheme reduces the determination of lengths to the measurements of times or frequencies, which can be measured much more accurately than lengths [180]. Recently, the direct comparison of optical frequencies with the Cs standard in the microwave region has become possible with broadband frequency combs generated by visible femtosecond lasers. These frequency combs represent equidistant frequencies, separated by about 100 MHz, which span a wide frequency range, typically over 10^{14} Hz. They allow absolute frequency measurements. This method will be discussed in Vol. 2, Sect. 14.7.

4.4.1 Precision and Accuracy of Wavelength Measurements

Resolving power and light-gathering power are not the only criteria by which a wavelength-dispersing instrument should be judged. A very important question is the attainable *precision* and *accuracy* of absolute wavelength measurements.

To measure a physical quantity means to *compare* it with a reference standard. This comparison involves statistical and systematic errors. Measuring the same quantity n times will yield values X_i that scatter around the mean value

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\ .$$

The attainable **precision** for such a set of measurements is determined by statistical errors and is mainly limited by the signal-to-noise ratio for a single measurement and by the number n of measurements (i.e., by the total measuring time). The precision can be characterized by the *standard deviation* [181, 182],

$$\sigma = \left(\sum_{i=1}^{n} \frac{(\overline{X} - X_i)^2}{n}\right)^{1/2}\ . \tag{4.106}$$

The adopted mean value $\overline{X}$, averaged over many measured values X_i, is claimed to have a certain *accuracy*, which is a measure of the reliability of this value, expressed by its probable deviation $\Delta\overline{X}$ from the unknown "true" value X. A stated accuracy of $\overline{X}/\Delta\overline{X}$ means a certain confidence that the true value X is within $\overline{X} \pm \Delta\overline{X}$. Since the accuracy is determined not only by statistical errors but, particularly, by systematic errors of the apparatus and measuring procedure, it is always lower than the precision. It is also influenced by the precision with which the reference standard can be measured and by the accuracy of its comparison with the value $\overline{X}$. Although the attainable accuracy depends on the experimental efforts and expenditures, the skill, imagination, and critical judgement of the experimentalist always have a major influence on the ultimate achieved and stated accuracy.

We shall characterize precision and accuracy by the relative uncertainties of the measured quantity X, expressed by the ratios

$$\frac{\sigma}{X} \quad \text{or} \quad \frac{\Delta\overline{X}}{X} \,,$$

respectively. A series of measurements with a standard deviation $\sigma = 10^{-8}\overline{X}$ has a relative uncertainty of 10^{-8} or a precision of 10^8. Often one says that the precision is 10^{-8}, although this statement has the disadvantage that a high precision is expressed by a small number.

Let us now briefly examine the attainable precision and accuracy of wavelength measurements with the different instruments discussed above. Although both quantities are correlated with the resolving power and the attainable signal-to-noise ratio, they are furthermore influenced by many other instrumental conditions, such as backlash of the monochromator drive, or asymmetric line profiles caused by imaging errors, or shrinking of the photographic film during the developing process. Without such additional error sources, the precision could be much higher than the resolving power, because the center of a symmetric line profile can be measured to a small fraction ϵ of the halfwidth. The value of ϵ depends on the attainable signal-to-noise ratio, which is determined, apart from other factors, by the étendue of the spectrometer. We see that for the precision of wavelength measurements, the product of resolving power R and étendue U, RU, discussed in the previous section, plays an important role.

For scanning monochromators with photoelectric recording, the main limitation for the attainable accuracy is the backlash of the grating-drive and nonuniformities of the gears, which limits the reliability of linear extrapolation between two calibration lines. Carefully designed monochromators have errors due to the drive that are less than $0.1\,\mathrm{cm}^{-1}$, allowing a relative uncertainty of 10^{-5} or an accuracy of about 10^5 in the visible range.

In absorption spectroscopy with a tunable laser, the accuracy of line positions is also limited by the nonuniform scan speed $\mathrm{d}\lambda/\mathrm{d}t$ of the laser (Sect. 5.6). One has to record reference wavelength marks simultaneously with the spectrum in order to correct for the nonuniformities of $\mathrm{d}\lambda/\mathrm{d}t$.

A serious source of error with scanning spectrometers or scanning lasers is the distortion of the line profile and the shift of the line center caused by the time constant of the recording device. If the time constant τ is comparable with the time $\Delta t = \Delta\lambda/v_{sc}$ needed to scan through the halfwidth $\Delta\lambda$ of the line profile (which depends on the spectral resolution), the line becomes broadened, the maximum decreases, and the center wavelength is shifted. The line shift $\delta\lambda$ depends on the scanning speed v_{sc} [nm/min] and is approximately $\delta\lambda = v_{sc}\tau = (d\lambda/dt)\tau$ [122].

Example 4.22
With a scanning speed $v_{sc} = 10\,\text{nm/min}$ and a time constant of the recorder $\tau = 1\,\text{s}$ the line shift is already $\delta\lambda = 0.15\,\text{nm}$!

Because of the additional line broadening, the resolving power is reduced. If this reduction is to be less than 10 %, the scanning speed must be below $v_{sc} < 0.24\Delta\lambda/\tau$. With $\Delta\lambda = 0.02\,\text{nm}$, $\tau = 1\,\text{s} \rightarrow v_{sc} < 0.3\,\text{nm/min}$.

Photographic recording avoids these problems and therefore allows a more accurate wavelength determination at the expense of an inconvenient developing process of the photoplate and the subsequent measuring procedure to determine the line positions. A typical figure for the standard deviation for a 3-m spectrograph is $0.01\,\text{cm}^{-1}$. Imaging errors causing curved lines, asymmetric line profiles due to misalignment, and backlash of the microdensitometer used for measuring the line positions on the photoplate are the main sources of errors.

Modern devices use photodiodes or CCD arrays (Sect. 4.5.2) instead of photoplates. With a diode width of 25 μm, the peak of a symmetric line profile extending over 3–5 diodes can be determined by a least-squares fit to a model profile within 1–5 μm, depending on the S/N ratio. When the array is placed behind a spectrometer with a dispersion of 1 mm/nm, the center of the line can be determined within 10^{-3} nm. Since the signals are read electronically, there are no moving parts in the device and any mechanical error source (backlash) is eliminated.

The highest accuray (i.e., the lowest uncertainty) can be achieved with modern *wavemeters*, which we shall discuss in Sect. 4.4.2.

4.4.2 Today's Wavemeters

The different types of wavemeters for very accurate measurements of laser wavelengths are based on modifications of the Michelson interferometer [184], the Fizeau interferometer [185], or on a combination of several Fabry–Perot interferometers with different free spectral ranges [186–188]. The wavelength is measured either by monitoring the spatial distribution of the interference pattern with photodiode arrays, or by using traveling devices with electronic counting of the interference

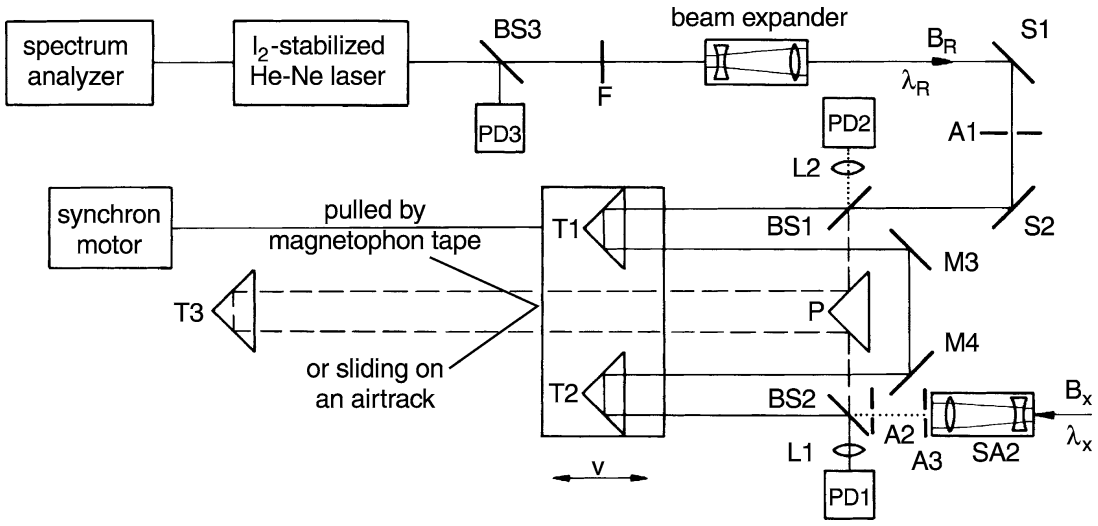

Figure 4.69 Traveling Michelson interferometer for accurate measurements of wavelengths of single-mode cw lasers

fringes. Nowadays several versions of wavemeters are commercially available which reach uncertainties of $\pm 0.2\,\mathrm{pm}$ (accuracies $\nu/\delta\nu$ of about 10^{+7}). They can operate over a wide spectral range from 300 nm to 5 μm.

a) The Michelson Wavemeter

Figure 4.69 illustrates the principle of a traveling-wave Michelson-type interferometer as used in our laboratory. Such a wavemeter was first demonstrated in a slightly different version by Hall and Lee [184] and by Kowalski et al. [190]. The beams B_R of a reference laser and B_x of a laser with unknown wavelength λ_x traverse the interferometer on identical paths, but in opposite directions. Both incoming beams are split into two partial beams by the beam splitters BS1 and BS2, respectively. One of the partial beams travels the constant path BS1–P–T3–P–BS2 for the reference beam, and in the opposite direction for the beam B_X. The second partial beam travels the variable path BS1–T1–M3–M4–T2–BS2 for B_R, and in the opposite direction for B_X. The moving corner-cube reflectors T1 and T2 are mounted on a carriage, which either travels with wheels on rods or slides on an airtrack.

The *corner-cube reflectors* guarantee that the incoming light beam is always reflected exactly parallel to its indicent direction, irrespective of slight misalignments or movements of the traveling reflector. The two partial beams (BS1–T1–M3–M4–T2–BS2 and BS1–P–T3–P–BS2) for the reference laser interfere at the detector PD1, and the two beams BS2–T2–M4–M3–T1–BS1 and BS2–P–T3–P–BS1 from the unknown laser interfere at the detector PD2. When the carriage is moving at a speed $v = \mathrm{d}x/\mathrm{d}t$ the phase difference $\delta(t)$ between the two interfering beams changes as

$$\delta(t) = 2\pi\frac{\Delta s}{\lambda} = 2\pi \cdot 4\frac{\mathrm{d}x}{\mathrm{d}t}\frac{t}{\lambda} = 8\pi\frac{vt}{\lambda}\,, \tag{4.107}$$

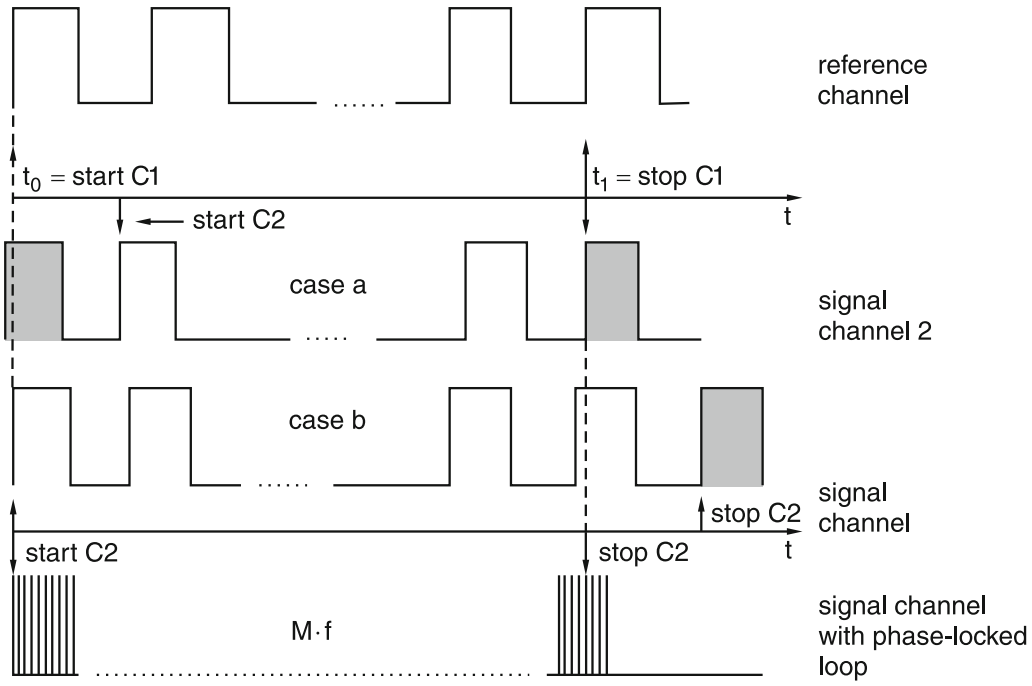

Figure 4.70 Signal sequences in the two detection channels of the traveling Michelson wavemeter. The grey signal pulses are not counted

where the factor 4 stems from the fact that the optical path difference Δs has been doubled by introducing two corner-cube reflectors. The rates of interference maxima, which occur for $\delta = m2\pi$, are counted by PD2 for the unknown wavelength λ_X and by PD1 for the reference wavelength λ_R. The unknown wavelength λ_X can be obtained from the ratio of both counting rates if proper corrections are made for the dispersion $n(\lambda_\mathrm{R}) - n(\lambda_X)$ of air. An electronic device produces a short voltage pulse each time the line-varying interference intensity passes through zero. These pulses are counted.

The signal lines to both counters are simultaneously opened at the time t_0 when the detector PD2 just delivers a trigger signal. Both counters are simultaneously stopped at the time t_1 when PD2 has reached the preset number N_0. From

$$\Delta t = t_1 - t_0 = N_0 \lambda_X / 4v = (N_\mathrm{R} + \epsilon)\lambda_\mathrm{R}/4v \ ,$$

we obtain for the vacuum wavelength λ_X^0

$$\lambda_X^0 = \frac{N_\mathrm{R} + \epsilon}{N_0} \lambda_\mathrm{R}^0 \frac{n(\lambda_X, P, T)}{n(\lambda_\mathrm{R}, P, T)} \ . \tag{4.108a}$$

The unknown fractional number $\epsilon < 2$ takes into account that the trigger signals from PD1, which define the start and stop times t_0 and t_1 (Fig. 4.69), may not exactly coincide with the pulse rise times in channel 2. The two worst cases are shown in Fig. 4.70. For case a, the trigger pulse at t_0 just misses the rise of the signal pulse, but the trigger at t_1 just coincides with the rise of a signal pulse. This means that the

signal channel counts one pulse less than it should. In case b, the start pulse at t_0 coincides with the rise time of a signal pulse, but the stop pulse just misses a signal pulse. In this case, the signal channel counts one pulse more than it should.

For a maximum optical path difference $\Delta s = 4\,\mathrm{m}$, the number of counts for $\lambda = 500\,\mathrm{nm}$ is 8×10^6, which allows a precision of about 10^7, if the counting error is not larger than 1. Provided the signal-to-noise ratio is sufficiently high, the attainable precision can, however, be enhanced by interpolations between two successive counts using a *phase-locked loop* [191, 192]. This is an electronic device that multiplies the frequency of the incoming signal by a factor M while always being locked to the phase of the incoming signal. Assume that the counting rate $f_\mathrm{R} = 4v/\lambda_\mathrm{R}$ in the reference channel is multiplied by M. Then the unknown wavelength λ_X is determined by

$$\lambda_X^0 = \frac{MN_\mathrm{R} + \epsilon}{MN_0}\lambda_\mathrm{R}^0 \frac{n_X}{n_\mathrm{R}} = \frac{N_\mathrm{R} + \varepsilon/M}{N_0}\lambda_\mathrm{R}^0 \frac{n_x}{n_\mathrm{R}} \,. \tag{4.108b}$$

For $M = 100$ the limitation of the accuracy by the counting error due to the unknown fractional number ϵ is reduced by a factor of 100.

Instead of the phase-locked loop a coincidence curcuit may be employed. Here the signal paths to both counters are opened and closed at selected times t_0 and t_1, when both trigger signals from PD2 and PD1 coincide within a small time interval, say 10^{-8} s. Both techniques reduce the counting uncertainty to a value below 2×10^{-9}.

In general, the attainable accuracy, however, is lower because it is influenced by several sources of systematic errors. One is a misalignment of the interferometer, which causes both beams to travel slightly different path lengths. Another point that has to be considered is the curvature of the wavefronts in the diffraction-limited Gaussian beams (Sect. 5.3). This curvature can be reduced by expanding the beams through telescopes (Fig. 4.69). The uncertainty of the reference wavelength λ_R and the accuracy of measuring the refractive index $n(\lambda)$ of air are further error sources.

The maximum relative uncertainty of the absolute vacuum wavelength λ_X can be written as a sum of five terms:

$$\left|\frac{\Delta\lambda_X}{\lambda_X}\right| \le \left|\frac{\Delta\lambda_\mathrm{R}}{\lambda_\mathrm{R}}\right| + \left|\frac{\epsilon}{MN_\mathrm{R}}\right| + \left|\frac{\Delta r}{r}\right| + \left|\frac{\delta s}{\Delta s}\right| + \left|\frac{\delta\phi}{2\pi N_0}\right| \,, \tag{4.109}$$

where $r = n(\lambda_X)/n(\lambda_\mathrm{R})$ is the ratio of the refractive indices, δs is the difference of the travel paths for reference and signal beams, and $\delta\phi$ is the phase front variation in the detector plane. Let us briefly estimate the magnitude of the different terms in (4.109):

- The wavelength λ_R of the I_2-stabilized HeNe laser is known within an uncertainty $|\Delta\lambda_\mathrm{R}/\lambda_\mathrm{R}| < 10^{-10}$ [180]. Its frequency stability is better than 100 kHz, i.e., $|\Delta\nu/\nu| < 2 \times 10^{-10}$. This means that the first term in (4.109) contributes at most 3×10^{-10} to the uncertainty of λ_X.
- With $\epsilon = 1.5$, $M = 100$, and $N_\mathrm{R} = 8 \times 10^6$, the second term is about 2×10^{-9}.

- The index of refraction, $n(\lambda, p, T)$, depends on the wavelength λ, on the total air pressure, on the partial pressures of H_2O and CO_2, and on the temperature. If the total pressure is measured within 0.5 mbar, the temperature T within 0.1 K, and the relative humidity within 5 %, the refractive index can be calculated from formulas given by Edlen [193] and Owens [194].
  With the stated accuracies, the third term in (4.109) becomes

$$|\Delta r/r| \approx 1 \times 10^{-3} \, |n_0(\lambda_X) - n_0(\lambda_R)| \ , \tag{4.110}$$

  where n_0 is the refractive index for dry air under standard conditions ($T_0 = 15\,°C$, $p_0 = 1013\,hPa$). The contribution of the third term depends on the wavelength difference $\Delta\lambda = \lambda_R - \lambda_X$. For $\Delta\lambda = 1\,nm$ one obtains $|\Delta r/r| < 10^{-11}$, while for $\Delta\lambda = 200\,nm$ this term becomes, with $|\Delta r/r| \approx 5 \times 10^{-9}$, a serious limitation of the accuracy of $|\Delta\lambda_X/\lambda_X|$.
- The magnitude of the fourth term $|\delta s/\Delta s|$ depends on the effort put into the alignment of the two laser beams within the interferometer. If the two beams are tilted against each other by a small angle α, the two path lengths for λ_X and λ_R differ by

$$\delta s = \Delta s(\lambda_R) - \Delta s(\lambda_X) = \Delta s_R(1 - \cos\alpha) \approx (\alpha^2/2)\Delta s_R \ .$$

  With $\alpha = 10^{-4}$ rad, the systematic relative error becomes

$$|\delta s/\Delta s| \approx 5 \times 10^{-9} \ .$$

  It is therefore necessary to align both beams very carefully.
- With a surface quality of $\lambda/10$ for all mirrors and beam splitters, the distortions of the wavefront are already visible in the interference pattern. However, plane waves are focused onto the detector area and the phase of the detector signal is due to an average over the cross section of the enlarged beam ($\approx 1\,cm^2$). This averaging minimizes the effect of wavefront distortion on the accuracy of λ_X. If the modulation of the interference intensity (4.37) exceeds 90 %, this term may be neglected.

With careful alignment, good optical quality of all optical surfaces and accurate recording of p, T, and P_{H_2O}, the total uncertainty of λ_X can be pushed below 10^{-8}. This gives an absolute uncertainty $\Delta\nu_x \approx 3\,MHz$ of the optical frequency $\nu_x = 5 \times 10^{14}\,s^{-1}$ for a wavelength separation between λ_R and λ_x of $\Delta\lambda \approx 120\,nm$. This has been proved by a comparison of independently measured wavelengths $\lambda_x = 514.5\,nm$ (I_2-stabilized argon laser) and $\lambda_R = 632.9\,nm$ (I_2-stabilized HeNe laser) [195].

When cw dye laser wavelengths are measured, another source of error arises. Due to air bubbles in the dye jet or dust particles within the resonator beam waist, the dye laser emission may be interrupted for a few microseconds. If this happens while counting the wavelength a few counts are missing. This can be avoided by using an additional phase-locked loop with a multiplication factor $M_x = 1$ in the

counting channel of PD_x. If the time constant of the phase-locked loop is larger than 10 µs, it continues to oscillate at the counting frequency during the few microseconds of dye laser beam interruptions.

There are several different designs of Michelson wavemeters that are commercially available and are described in [197–199].

b) Sigmameter

While the traveling Michelson is restricted to cw lasers, a motionless Michelson interferometer was designed by Jacquinot, et al. [200], which includes no moving parts and can be used for cw as well as for pulsed lasers. Figure 4.71 illustrates

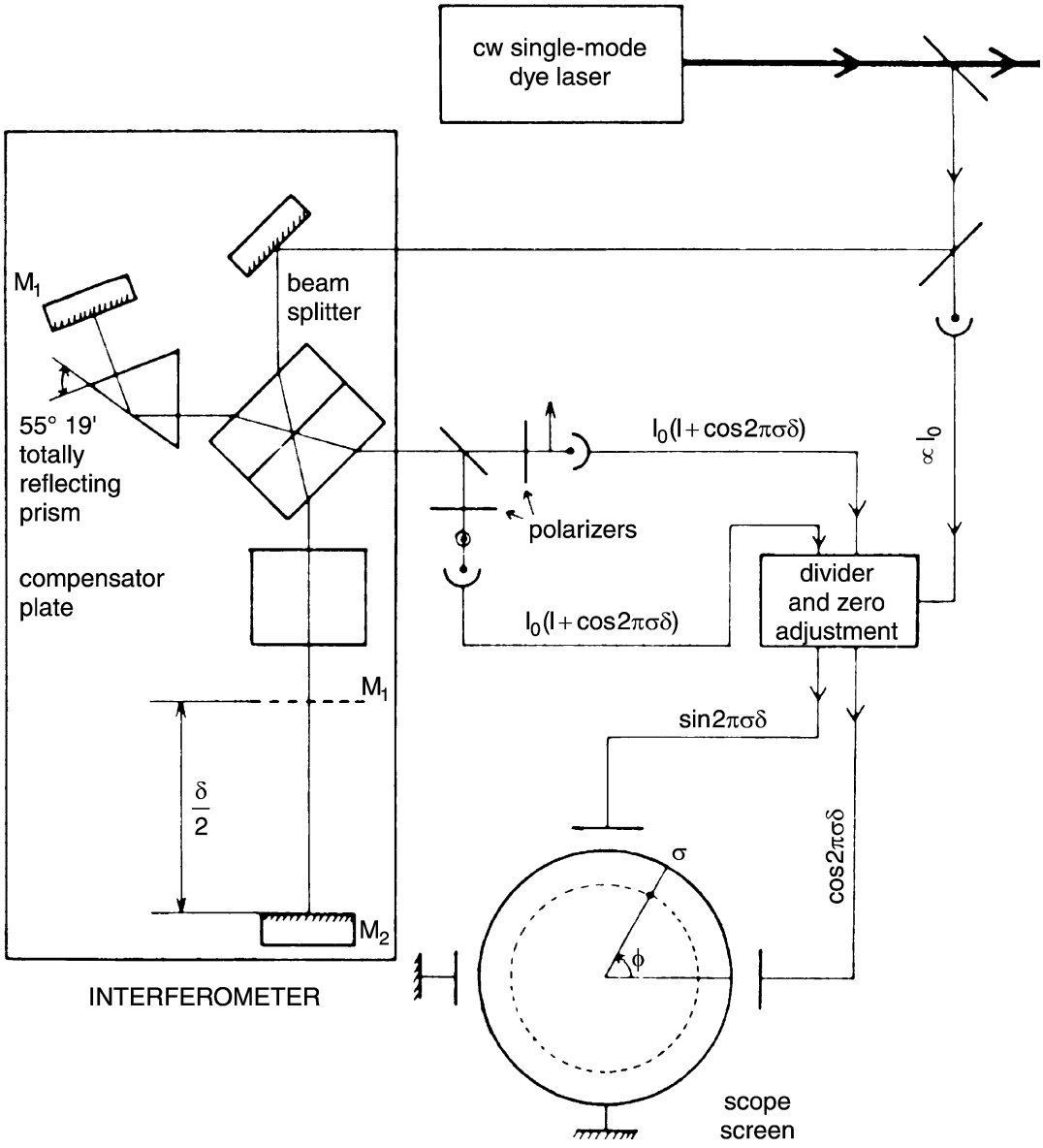

Figure 4.71 Sigmameter [200]

its operation. The basic element is a Michelson interferometer with a *fixed* path difference δ. The laser beam enters the interferometer polarized at 45° with respect to the plane of Fig. 4.71. When inserting a prism into one arm of the interferometer, where the beam is totally reflected at the prism base, a phase difference $\Delta\varphi$ is introduced between the two components polarized parallel and perpendicular to the totally reflecting surface. The value of $\Delta\varphi$ depends, according to Fresnel's formulas [129], on the incidence angle α and can be made $\pi/2$ for $\alpha = 55°19'$ and $n = 1.52$. The interference signal at the exit of the interferometer is recorded separately for the two polarizations and one obtains, because of the phase shifts $\pi/2$, $I_{||} = I_0(1 + \cos 2\pi\delta/\lambda)$ and $I_\perp = I_0(1 + \sin 2\pi\delta/\lambda)$. From these signals it is possible to deduce the wave number $\sigma = 1/\lambda$ modulo $1/\delta$, since all wave numbers $\sigma_m = \sigma_0 + m/\delta$ $(m = 1, 2, 3, \ldots)$ give the same interference signals. Using several interferometers of the same type with a common mirror M1 but different positions of M2, which have path differences in geometric ratios, such as 50 cm, 5 cm, 0.5 cm, and 0.05 cm, the wave number σ can be deduced unambiguously with an accuracy determined by the interferometer with the highest path difference. The actual path differences δ_i are calibrated with a reference line and are servo-locked to this line. The precision obtained with this instrument is about 5 MHz, which is comparable with that of the traveling Michelson interferometer. The measuring time, however, is much less since the different δ_i can be determined simultaneously. This instrument is more difficult to build but easier to handle. Since it measures wave numbers $\sigma = 1/\lambda$, the inventors called it a *sigmameter*.

c) Computer-Controlled Fabry–Perot Wavemeter

Another approach to accurate wavelength measurements of pulsed and cw lasers, which can be also applied to incoherent sources, relies on a combination of a small grating monochromator and three Fabry–Perot etalons [186–188]. The incoming laser beam is sent simultaneously through the monochromator and three temperature-stabilized Fabry–Perot interferometers with different free spectral ranges $\delta\nu_i$ (Fig. 4.72). In order to match the laser beam profile to the sensitive area of the linear diode arrays (25 mm × 50 µm), focusing with cylindrical lenses Z_i is utilized. The divergence of the beams in the plane of Fig. 4.72 is optimized by the spherical lenses L_i in such a way that the diode arrays detect 4–6 FPI fringes (Fig. 4.73). The linear arrays have to be properly aligned so that they coincide with a diameter through the center of the ring system. According to (4.72), the wavelength λ can be determined from the ring diameters D_p and the excess ϵ, provided the integer order m_0 is known, which means that λ must already be known at least within one-half of a free spectral range (Sect. 4.3).

The device is calibrated with different lines from a cw dye laser that are simultaneously measured with the traveling Michelson wavemeter (see above). This calibration allows:

- The unambiguous correlation between wavelength λ and the position of the illuminated diode of array 1 behind the monochromator with an accuracy of

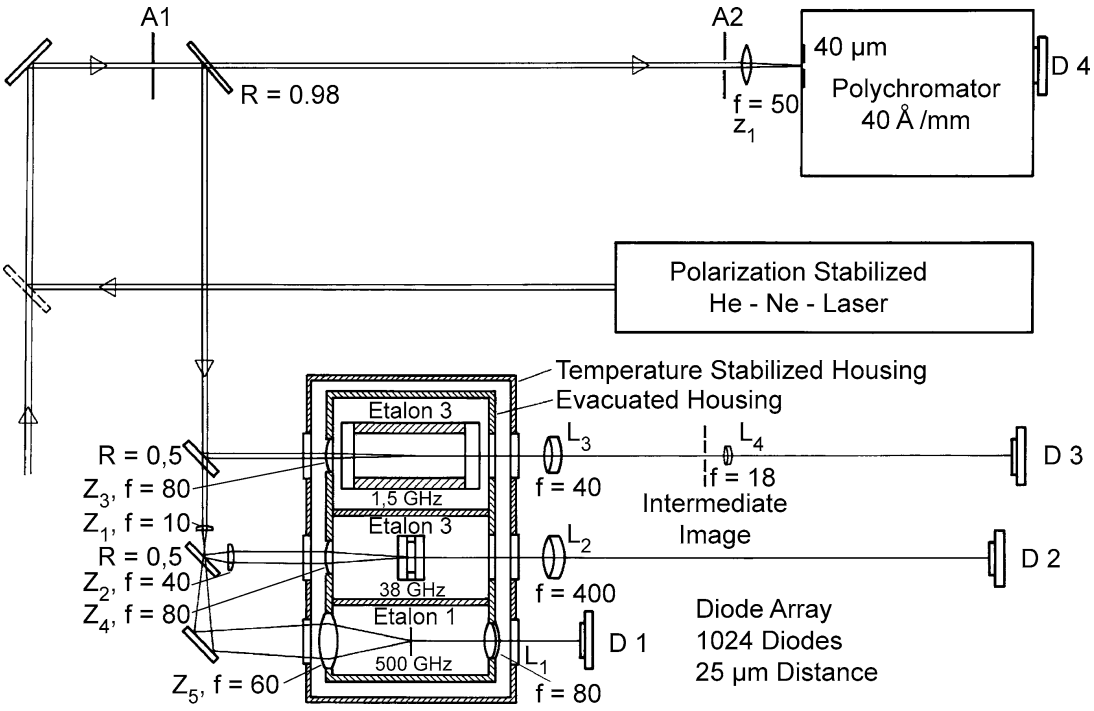

Figure 4.72 Wavemeter for pulsed and cw lasers, based on a combination of a small polychromator and three FPI with widely differing free spectral ranges [195]

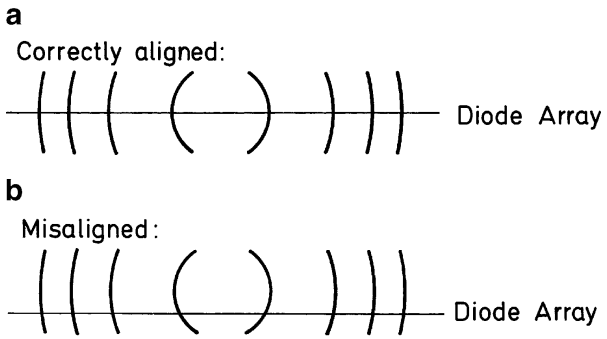

Figure 4.73 Measuring interference ring diameters with a linear diode array: **a** correct alignment; **b** misaligned diode array

± 0.1 nm, which is sufficient to determine λ within 0.5 of the free spectral range of etalon 1;

- The accurate determination of nd for all three FPI.

If the free spectral range $\delta\nu_1$ of the thin FPI is at least twice as large as the uncertainty $\Delta\nu$ of the monochromator measurement, the integer order m_0 of FPI1 can be unambiguously determined. The measurement of the ring diameters improves the accuracy by a factor of about 20. This is sufficient to determine the larger integer order m_0 of FPI2; from its ring diameters, λ can be measured with an accuracy 20 times higher than that from FPI1. The final wavelength determination uses the ring diameters of the large FPI3. Its accuracy reaches about 1 % of the free spectral range of FPI3.

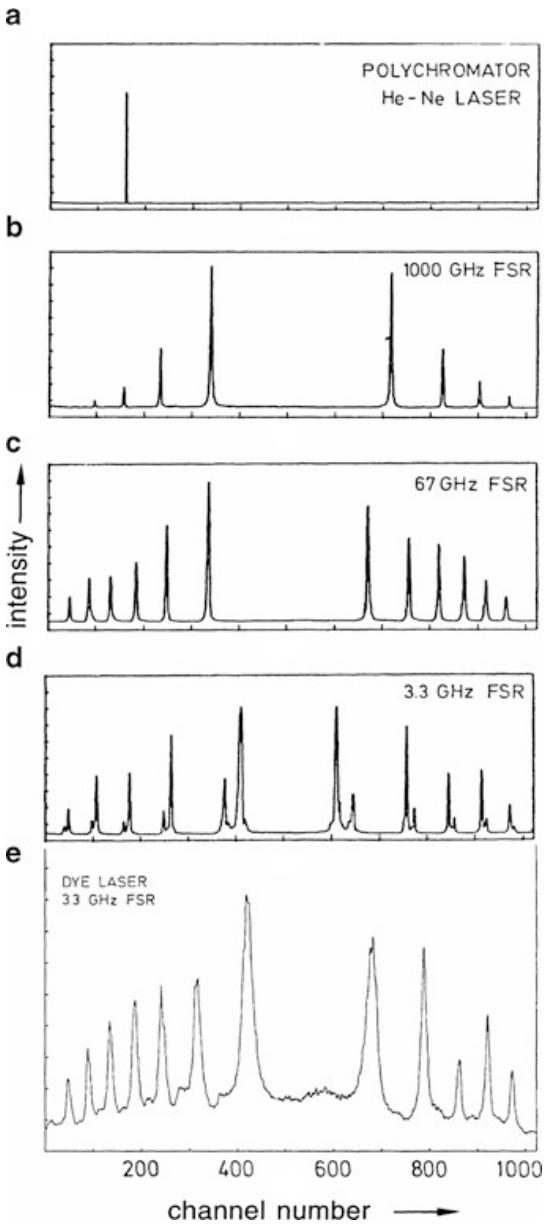

Figure 4.74 Output signals at the polychromator and the three diode arrays of the FPI wavemeter, which had been illuminated by a cw HeNe laser oscillating on two axial modes (**a**–**d**). The lowest figure shows the ring intensity pattern of an excimer-pumped single-mode dye laser measured behind a FPI with 3.3 GHz free spectral range [187]

The whole measuring cycle is controlled by a computer. For pulsed lasers, one pulse (with an energy of $\geq 5\,\mu J$) is sufficient to initiate the device, while for cw lasers, a few microwatts input power are sufficient. The arrays are read out by the computer and the signals can be displayed on a screen. Such signals for the arrays D1–D4 are shown in Fig. 4.74 for a HeNe laser oscillating on two longitudinal modes and for a pulsed dye laser.

Since the optical distances $n_i d_i$ of the FPI depend critically on temperature and pressure, all FPI must be kept in a temperature-stabilized pressure-tight box. Furthermore, a stabilized HeNe laser can be used to control long-term drift of the FPI [195].

Example 4.23
With a free spectral range of $\delta\nu = 1\,\text{GHz}$, the uncertainty of calibration and of the determination of an unknown wavelength are both about 10 MHz. This gives an absolute uncertainty of less than 20 MHz. For the optical frequency $\nu = 6 \times 10^{14}\,\text{Hz}$, the relative accuracy is then $\Delta\nu/\nu \leq 3 \times 10^{-8}$.

d) Fizeau Wavemeter

The Fizeau wavemeter constructed by Snyder [201] can be used for pulsed and cw lasers. While its optical design is simpler than that of the sigmameter and the FPI wavemeter, its accuracy is slightly lower. Its basic principle is shown in Fig. 4.75b. The incident laser beam is focused by an achromatic microscope lens system onto a small pinhole, which represents a nearly pointlike light source. The divergent light is transformed by a parabolic mirror into an enlarged parallel beam, which hits the Fizeau interferometer (FI) under an incident angle α (Fig. 4.75a). The FI consists of two fused quartz plates with a slightly wedged air gap ($\phi \approx 1/20°$). For small wedge angles ϕ, the optical path difference Δs between the constructively interfering beams 1 and $1'$ is approximately equal to that of a plane-parallel plate according to (4.48a), namely

$$\Delta s_1 = 2n\,d(z_1)\cos\beta = m\lambda\ .$$

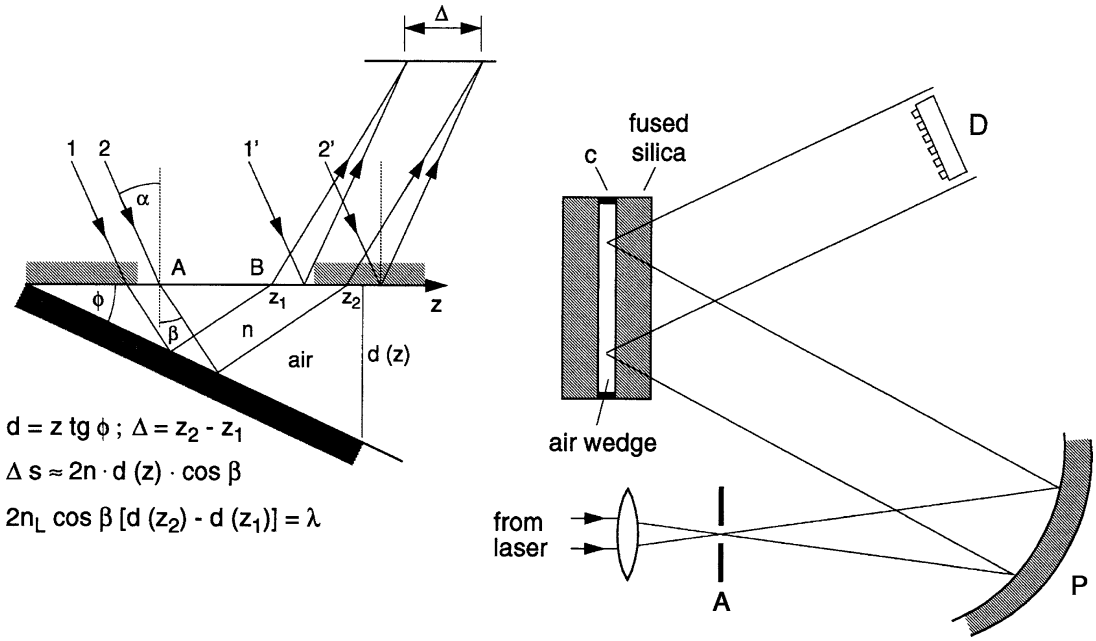

Figure 4.75 Fizeau wavemeter: **a** interference at a wedge (the wedge angle ϕ is greatly exaggerated); **b** schematic design; A, aperture as spatial filter; P, parabolic mirror; C, distance holder of cerodur; D, diode array

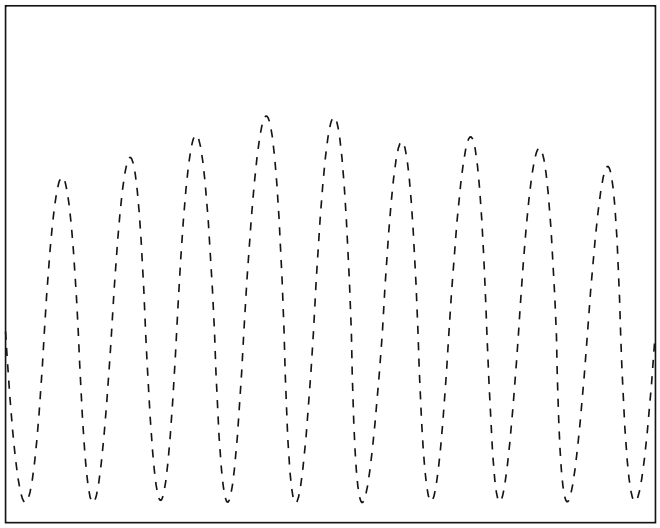

Figure 4.76 Densitometer trace of the fringe pattern in a Fizeau wavemeter [185]

The path difference between the beams 2 and 2′, which belong to the next interference order, is $\Delta s_2 = (m+1)\lambda$. The interference of the reflected light produces a pattern of parallel fringes (Fig. 4.76) with the separation

$$\Delta = z_2 - z_1 = \frac{d(z_2) - d(z_1)}{\tan\phi} = \frac{\lambda}{2n\tan\phi\cos\beta}\,, \tag{4.111}$$

which depends on the wavelength λ, the wedge angle ϕ, the angle of incidence α, and the refractive index n of air.

Changing the wavelength λ causes a shift Δz of the fringe pattern and a slight change of the fringe separation Δ. For a change of λ by one free spectral range

$$\delta\lambda = \frac{\lambda^2}{2nd\cos\beta}\,. \tag{4.112}$$

and Δz is equal to the fringe separation Δ. Therefore the two fringe patterns for λ and $\lambda + \delta\lambda$ look identical, apart from the slight change of Δ. It is therefore essential to know λ at least within $\pm\delta\lambda/2$. This is possible from a measurement of Δ. With a diode array of 1024 diodes, the fringe separation Δ can be obtained from a least-squares fit to the measured intensity distribution $I(z)$ with a relative accuracy of 10^{-4}, which yields an absolute value of λ within $\pm 10^{-4}\lambda$ [202].

With a value $d = 1\,\text{mm}$ of the air gap, the order of interference m is about 3000 at $\lambda = 500\,\text{nm}$. An accuracy of 10^{-4} is therefore sufficient for the unambiguous determination of m. Since the *position* of the interference fringes can be measured within 0.3 % of the fringe separation, the wavelength λ can be obtained within 0.3 % of a free spectral range, which gives the accuracy $\lambda/\Delta\lambda \approx 10^7$. The preliminary value of λ, deduced from the fringe separation Δ, and the final value, determined from the *fringe position*, are both obtained from the same FI after having calibrated the system with lines of known wavelengths.

The advantage of the Fizeau wavemeter is its compact design and its low price. A very elegant construction by Gardner [203, 204] is sketched in Fig. 4.77. The

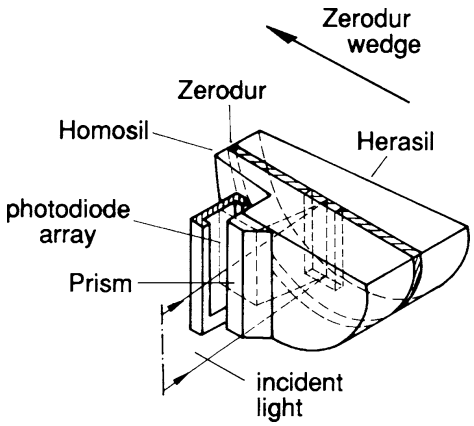

Figure 4.77 Compact design of a Fizeau wavemeter [203]

wedge air gap is fixed by a Zerodur spacer between the two interferometer plates and forms a pressure tight volume. Variations of air pressure in the surroundings therefore do not cause changes of n within the air gap. The reflected light is sent to the diode array by a totally reflecting prism. The data are processed by a small computer [205].

4.5 Detection of Light

For many applications in spectroscopy the sensitive detection of light and the accurate measurement of its intensity are of crucial importance for the successful performance of an experiment. The selection of the proper detector for optimum *sensitivity* and *accuracy* for the detection of radiation must take into account the following characteristic properties, which may differ for the various detector types:

- The spectral relative response $R(\lambda)$ of the detector, which determines the wavelength range in which the detector can be used. The knowledge of $R(\lambda)$ is essential for the comparison of the true relative intensities $I(\lambda_1)$ and $I(\lambda_2)$ at different wavelengths.
- The absolute sensitivity $S(\lambda) = V_s/P$, which is defined as the ratio of output signal V_s to incident radiation power P. If the output is a voltage, as in photovoltaic devices or in thermocouples, the sensitivity is expressed in units of volts per watt. In the case of photocurrent devices, such as photomultipliers, $S(\lambda)$ is given in amperes per watt. With the detector area A the sensitivity S can be expressed in terms of the irradiance I:

$$S(\lambda) = V_s/(AI) \; . \tag{4.113}$$

- The achievable signal-to-noise ratio V_s/V_n, which is, in principle, limited by the noise of the incident radiation. It may, in practice, be further reduced by inherent noise of the detector. The detector noise is often expressed by the *noise equivalent input power* (NEP), which means an incident radiation power that

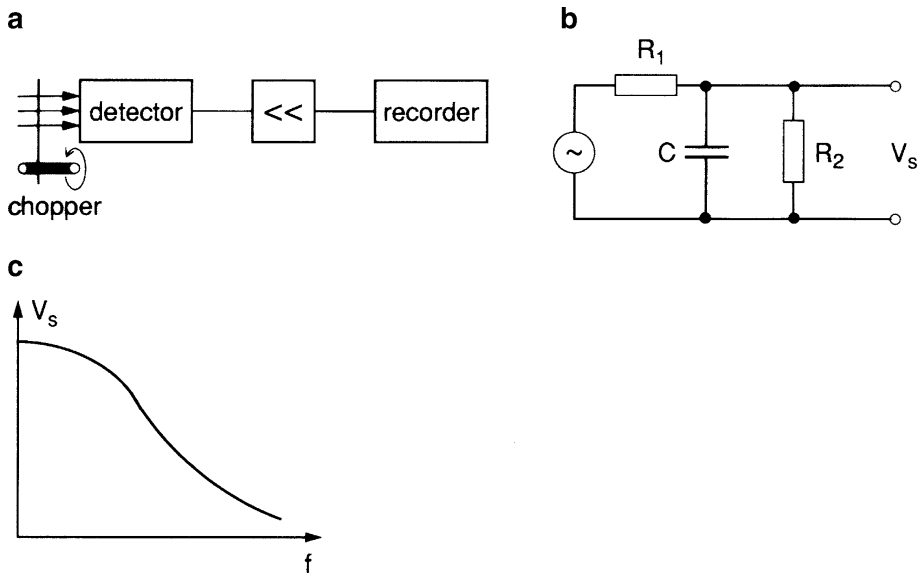

Figure 4.78 Typical detector: **a** schematic setup; **b** equivalent electrical circuit; **c** frequency response $V_s(f)$

generates the same output signal as the detector noise itself, thus yielding the signal-to-noise ratio $S/N = 1$. In infrared physics a figure of merit for the infrared detector is the detectivity

$$D = \frac{\sqrt{A\Delta f}}{P}\frac{V_s}{V_n} = \frac{\sqrt{A\Delta f}}{\text{NEP}} \,. \tag{4.114}$$

The specific detectivity D^* [cm s$^{-1/2}$ W^{-1}] gives the obtainable signalto-noise ratio V_s/V_n of a detector with the sensitive area $A = 1\,\text{cm}^2$ and the detector bandwidth $\Delta f = 1\,\text{Hz}$, at an incident radiation power of $P = 1\,\text{W}$. Because the noise equivalent input power is $\text{NEP} = P \cdot V_n/V_s$, the specific detectivity of a detector with the area 1 cm^2 and a bandwidth of 1 Hz is $D^* = 1/\text{NEP}$.

- The maximum intensity range in which the detector response is linear. It means that the output signal V_s is proportional to the incident radiation power P. This point is particularly important for applications where a wide range of intensities is covered. Examples are output-power measurements of pulsed lasers, Raman spectroscopy, and spectroscopic investigations of line broadening, when the intensities in the line wings may be many orders of magnitude smaller than at the center.
- The time or frequency response of the detector, characterized by its time constant τ. Many detectors show a frequency response that can be described by the model of a capacitor, which is charged through a resistor R_1 and discharged through R_2 (Fig. 4.78b). When a very short light pulse falls onto the detector, its output pulse is smeared out. If the output is a current $i(t)$ that is proportional to the incident radiation power $P(t)$ (as, for example, in photomultipliers), the output capacitance C is charged by this current and shows a voltage rise and fall,

determined by

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{1}{C}\left[i(t) - \frac{V}{R_2}\right] . \tag{4.115}$$

If the current pulse $i(t)$ lasts for the time T, the voltage $V(t)$ at the capacitor increases up to $t = T$ and for $R_2C \gg T$ reaches the peak voltage

$$V_{\max} = \frac{1}{C}\int_0^T i(t)\mathrm{d}t ,$$

which is determined by C and not by R_2! After the time T the voltage decays exponentially with the time constant $\tau = CR_2$. Therefore, the value of R_2 limits the repetition frequency f of pulses to $f < (R_2C)^{-1}$.
The time constant $\tau_1 = R_1C$ of the detector causes the output signal to rise slower than the incident input pulse and the time constant $\tau_2 = R_2C$ causes a slower decay than that of the input pulse. The time constants can be determined by modulating the continuous input radiation at the frequency f. The output signal of such a device is characterized by (see Example 4.14)

$$V_s(f) = \frac{V_s(0)}{\sqrt{1 + (2\pi f\tau)^2}} , \tag{4.116}$$

where $\tau = CR_1R_2/(R_1 + R_2) = \tau_1 \cdot \tau_2/(\tau_1 + \tau_2)$. At the modulation frequency $f = 1/(2\pi\tau)$, the output signal has decreased to $1/\sqrt{2}$ of its dc value. The knowledge of the detector time constant τ is essential for all applications where fast transient phenomena are to be monitored, such as atomic lifetimes or the time dependence of fast laser pulses (Vol. 2, Chap. 6).

- The price of a detector is another factor that cannot be ignored, since unfortunately it often restricts the optimum choice.

In this section we briefly discuss some detectors that are commonly used in laser spectroscopy. The different types can be divided into two categories, *thermal detectors* and *direct photodetectors*. In thermal detectors, the energy absorbed from the incident radiation raises the temperature and causes changes in the temperature-dependent properties of the detector, which can be monitored. Direct photodetectors are based either on the emission of photoelectrons from photocathodes, or on changes of the conductivity of semiconductors due to incident radiation, or on photovoltaic devices where a voltage is generated by the internal photoeffect. Whereas thermal detectors have a *wavelength-independent* sensitivity, photodetectors show a spectral response that depends on the work function of the emitting surface or on the band gap in semiconductors.

During recent years the development of image intensifiers, image converters, CCD cameras, and vidicon detectors has made impressive progress. At first pushed by military demands, these devices are now coming into use for light detection at

low levels, e.g., in Raman spectroscopy, or for monitoring the faint fluorescence of spurious molecular constituents. Because of their increasing importance we give a short survey of the principles of these devices and their application in laser spectroscopy. In time-resolved spectroscopy, subnanosecond detection can now be performed with fast phototubes in connection with transient digitizers, which resolve time intervals of less than 100 ps. Since such time-resolved experiments in laser spectroscopy with streak cameras and correlation techniques are discussed in Vol. 2, Chap. 6, we confine ourselves here to discussing only some of these modern devices from the point of view of spectroscopic instrumentation. A more extensive treatment of the characteristics and the performance of various detectors can be found in special monographs on detectors [206, 208, 211–217]. For reviews on photodetection techniques relevant in laser physics, see also [218–221].

4.5.1 Thermal Detectors

Because of their wavelength-independent sensitivity, thermal detectors are useful for calibration purposes, e.g., for an absolute measurement of the radiation power of cw lasers, or of the output energy of pulsed lasers. In the rugged form of medium-sensitivity calibrated calorimeters, they are convenient devices for any laser laboratory. With more sophisticated and delicate design, they have been developed as sensitive detectors for the whole spectral range, particularly for the infrared region, where other sensitive detectors are less abundant than in the visible range.

For a simple estimate of the sensitivity and its dependence on the detector parameters, such as the heat capacitance and thermal losses, we shall consider the following model [222]. Assume that the fraction β of the incident radiation power P is absorbed by a thermal detector with heat capacity H, which is connected to a heat sink at constant temperature T_s (Fig. 4.79a). When G is the thermal conductivity of the link between the detector and the heat sink, the temperature T of the detector under illumination can be obtained from

$$\beta P = H\frac{\mathrm{d}T}{\mathrm{d}t} + G(T - T_s) \; . \tag{4.117}$$

If the time-independent radiation power P_0 is switched on at $t = 0$, the time-dependent solution of (4.117) is

$$T = T_s + \frac{\beta P_0}{G}(1 - \mathrm{e}^{-(G/H)t}) \; . \tag{4.118}$$

The temperature T rises from the initial value T_s at $t = 0$ to the temperature $T = T_s + \Delta T$ for $t = \infty$ with the time constant $\tau = H/G$. The temperature rise for $t = \infty$

$$\Delta T = \frac{\beta P_0}{G} \tag{4.119}$$

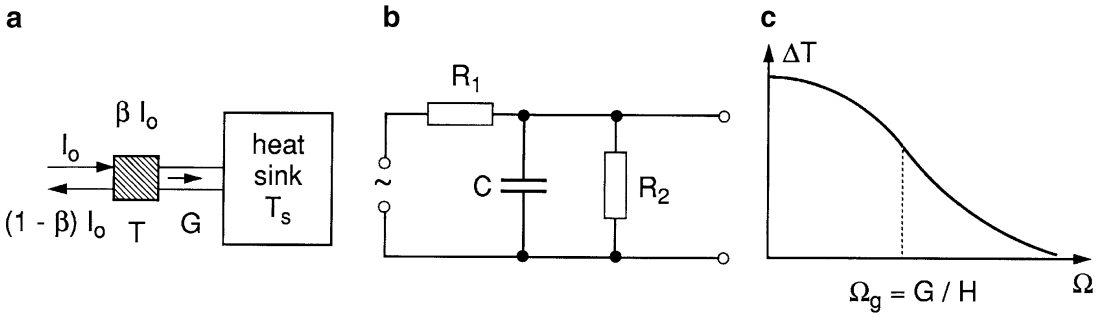

Figure 4.79 Model of a thermal detector: **a** schematic diagram; **b** equivalent electrical circuit; **c** frequency response $\Delta T(\Omega)$

is inversely proportional to the thermal losses G and does not depend on the heat capacity H, while the time constant of the rise $\tau = H/G$ depends on the ratio of both quantities. Small values of G make a thermal detector sensitive, but slow! It is therefore essential to realize small values of both quantities (H and G).

In general, P will be time dependent. When we assume the periodic function

$$P = P_0(1 + a\cos\Omega t) \,, \quad |a| \leq 1 \,, \tag{4.120}$$

we obtain, inserting (4.120) into (4.117), a detector temperature of

$$T(\Omega) = T_\mathrm{s} + \Delta T(1 + \cos(\Omega t + \varphi)) \,, \tag{4.121}$$

which depends on the modulation frequency Ω, and which shows a phase lag ϕ determined by

$$\tan\phi = \Omega H/G = \Omega\tau \,, \tag{4.122a}$$

and a modulation amplitude

$$\Delta T = \frac{a\beta P_0}{\sqrt{G^2 + \Omega^2 H^2}} = \frac{a\beta P_0}{G\sqrt{1 + \Omega^2\tau^2}} \,. \tag{4.122b}$$

At the frequency $\Omega_g = G/H = 1/\tau$, the amplitude ΔT decreases by a factor of $\sqrt{2}$ compared to its DC value.

Note The problem is equivalent to the analogous case of charging a capacitor ($C \leftrightarrow H$) through a resistor R_1 that discharges through R_2 ($R_2 \leftrightarrow 1/G$) (the charging current i corresponds to the radiation power P). The ratio $\tau = H/G$ ($H/G \leftrightarrow R_2C$) determines the time constant of the device (Fig. 4.79b).

We learn from (4.122b) that the sensitivity $S = \Delta T/P_0$ becomes large if G and H are made as small as possible. For modulation frequencies $\Omega > G/H$,

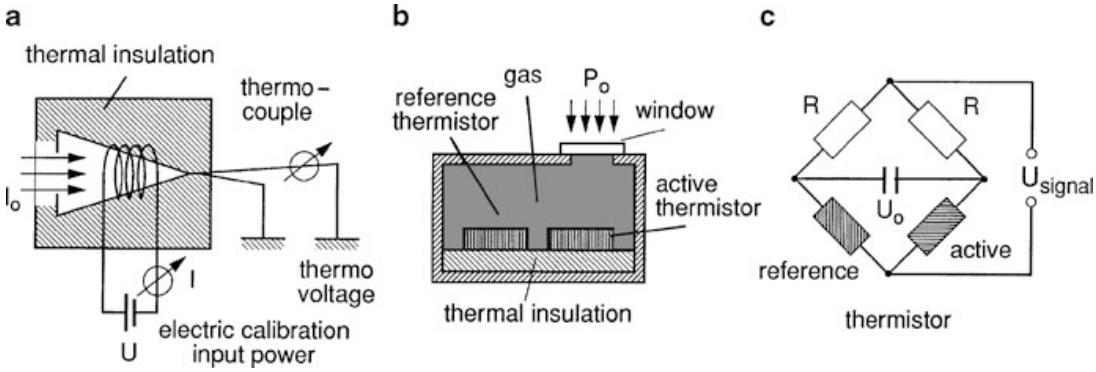

Figure 4.80 Calorimeter for measuring the output power of cw lasers or the output energy of pulsed lasers: **a** experimental design; **b** calorimeter with active irradiated thermistor and nonirradiated reference thermistor; **c** balanced bridge circuit

the amplitude ΔT will decrease approximately inversely to Ω. Since the time constant $\tau = H/G$ limits the frequency response of the detector, *a fast and sensitive detector should have a minimum heat capacity* H.

Since the specific heat decreases with decreasing temperature, thermal detectors with fast response but still high sensitivity, i.e. large value of G, should be operated at low temperatures.

For the calibration of the output power from cw lasers, the demand for high sensitivity is not as relevant since, in general, sufficiently large radiation power is available. Figure 4.80 depicts a simple home-made calorimeter and its circuit diagram. The radiation falls through a hole into a metal cone with a black inner surface. Because of the many reflections, the light has only a small chance of leaving the cone, ensuring that all light is absorbed. The absorbed power heats a thermocouple or a temperaturedependent resistor (thermistor) embedded in the cone. For calibration purposes, the cone can be heated by an electric wire. If the detector represents one part of a bridge (Fig. 4.80c) that is balanced for the electric input $W = UI$, but without incident radiation, the heating power has to be reduced by $\Delta W = P$ to maintain the balance with the incident radiation power P.

A system with higher accuracy uses the difference in output signals of two identical cones, where only one is irradiated (Fig. 4.80b).

For the measurement of output energies from pulsed lasers, the calorimeter should integrate the absorbed power at least over the pulse duration. From (4.117) we obtain

$$\int_0^{t_0} \beta P \, \mathrm{d}t = H\Delta T + \int_0^{t_0} G(T - T_s)\mathrm{d}t \; . \tag{4.123}$$

When the detector is thermally isolated, the heat conductivity G is small, therefore the second term may be completely neglected for sufficiently short pulse dura-

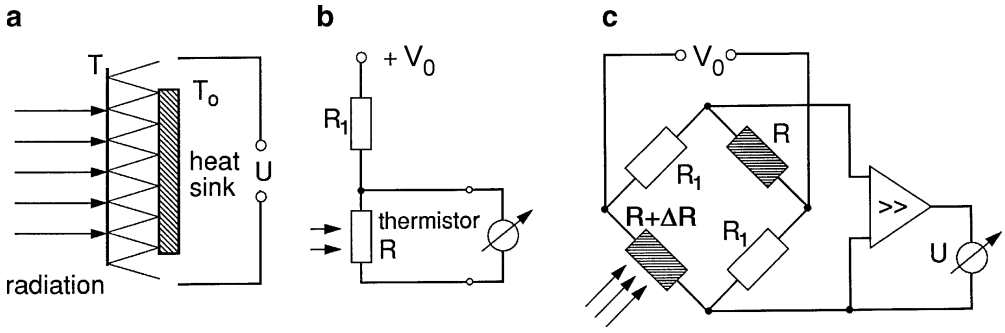

Figure 4.81 Schematic circuit diagram of a bolometer: **a** thermopile; **b** thermistor; and **c** bridge circuit with difference amplifier

tions t_0. The temperature rise

$$\Delta T = \frac{1}{H} \int_0^{t_0} \beta P \, \mathrm{d}t \ , \tag{4.124}$$

is then directly proportional to the input energy. Instead of the cw electric input for calibration (Fig. 4.80a), now a charged capacitor C is discharged through the heating coil. If the discharge time is matched to the laser pulse time, the heat conduction is the same for both cases and does not enter into calibration. If the temperature rise caused by the discharge of the capacitor equals that caused by the laser pulse, the pulse energy is $\frac{1}{2} C U^2$.

For more sensitive detection of low incident powers, *bolometers* and *Golay cells* are used. A special design for a bolometer consists of N thermocouples in series, where one junction touches the backside of a thin electrically insulating foil that is exposed to the incident radiation (Fig. 4.81a). The other junction is in contact with a heat sink. The output voltage is

$$U = N \frac{\mathrm{d}U}{\mathrm{d}T} \Delta T \ ,$$

where $\mathrm{d}U/\mathrm{d}T$ is the sensitivity of a single thermocouple.

Another version utilizes a thermistor that consists of a material with a large temperature coefficient $\alpha = (\mathrm{d}R/\mathrm{d}T)/R$ of the electrical resistance R. If a constant current i is fed through R (Fig. 4.81b), the incident power P that causes a temperature increase ΔT produces the voltage output signal

$$\Delta U = i \Delta R = i R \alpha \Delta T = \frac{V_0 R}{R + R_1} \alpha \Delta T \ , \tag{4.125}$$

where ΔT is determined from (4.121) as $\Delta T = \beta P (G^2 + \Omega^2 H^2)^{-1/2}$. The response $\Delta U / P$ of the detector is therefore proportional to i, R, and α, and decreases

with increasing H and G. At a constant supply voltage V_0, the current change Δi caused by the irradiation is, for $\Delta R \ll R + R_1$,

$$\Delta i = V_0 \left(\frac{1}{R_1 + R} - \frac{1}{R_1 + R + \Delta R} \right) \approx V_0 \frac{\Delta R}{(R_1 + R)^2} , \tag{4.126}$$

and can be generally neglected.

Since the input impedance of the following amplifier has to be larger than R, this puts an upper limit on R. Because any fluctuation of i causes a noise signal, the current i through the bolometer has to be extremely constant. This and the fact that the temperature rise due to Joule's heating should be small, limits the maximum current through the bolometer.

Equations (4.125 and 4.121) demonstrate again that small values of G and H are desirable. Even with perfect thermal isolation, heat radiation is still present and limits the lower value of G. At the temperature difference ΔT between a bolometer and its surroundings, the Stefan–Boltzmann law gives for the net radiation flux ΔP to the surroundings from the detector with the emitting area A^* and the emissivity $\epsilon \leq 1$

$$\Delta P = 4A\epsilon\sigma T^3 \Delta T , \tag{4.127}$$

where $\sigma = 5.77 \times 10^{-8}\,\mathrm{W/m^2\,K^{-4}}$ is the Stefan–Boltzmann constant. The minimum thermal conductivity is therefore

$$G_\mathrm{m} = 4A\sigma\epsilon T^3 , \tag{4.128}$$

even for the ideal case where no other heat links to the surroundings exist. This limits the detection sensitivity to a minimum input radiation of about $10^{-10}\,\mathrm{W}$ for detectors operating at room temperatures and with a bandwidth of 1 Hz. It is therefore advantageous to cool the bolometer, which furthermore decreases the heat capacity.

This cooling has the additional advantage that the slope of the function $\mathrm{d}R/\mathrm{d}T$ becomes larger at low temperatures T. Two different materials can be utilized, as discussed below.

In semiconductors the electrical conductivity is proportional to the electron density n_e in the conduction band. With the band gap ΔE_G this density is, according to the Boltzmann relation

$$\frac{n_\mathrm{e}(T)}{n_\mathrm{e}(T + \Delta T)} = \exp\left(-\frac{\Delta E_\mathrm{G} \Delta T}{2kT^2} \right) , \tag{4.129}$$

and is very sensitively dependent on temperature.

The quantity $\mathrm{d}R/\mathrm{d}T$ becomes exceedingly large at the critical temperature T_c of superconducting materials. If the bolometer is always kept at this temperature T_c by a temperature control, the incident radiation power P can be very sensitively measured by the magnitude of the feedback control signal used to compensate for the absorbed radiation power [224–226].

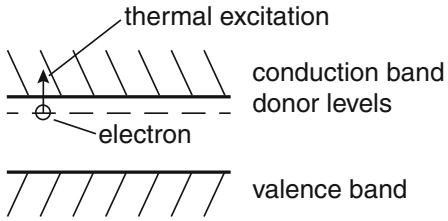

Figure 4.82 Thermal excitation of electrons from donor levels into the conduction band

Example 4.24
With $\int P\,\mathrm{d}t = 10^{-12}\,\mathrm{Ws}$, $\beta = 1$, $H = 10^{-11}\,\mathrm{Ws/K}$ we obtain from (4.124): $\Delta T = 0.1\,\mathrm{K}$. With $\alpha = 10^{-4}/\mathrm{K}$ and $R = 10\,\Omega$, $R_1 = 10\,\Omega$, $V_0 = 1\,\mathrm{V}$, the current change is $\Delta i = 2.5 \times 10^{-6}\,\mathrm{A}$ and the voltage change is $\Delta V = R\Delta i = 2.5 \times 10^{-5}\,\mathrm{V}$, which is readily detected.

Another material used for sensitive bolometers is a thin small disc of doped silicon, where the dopants are donor atoms with energy levels slightly below the conduction band (Fig. 4.82). A small temperature rise ΔT increases the fraction of ionized donors exponentially, thus producing free electrons in the conduction band. Such bolometers have to be operated at low temperatures in order to increase their sensitivity. The detectivity D^* (4.114) increases with falling temperature because the noise decreases (Fig. 4.83).

For thermal detectors the heat conduction G limits the sensitivity and the heat capacity H the frequency response. Since the specific heat decreases with the temperature, low temperatures improve both the sensitivity and the frequency response.

In Fig. 4.84 the whole setup for a bolometer operated at liquid helium temperatures is shown, including the liquid nitrogen and helium containers. Pumping the evaporating helium gas away drops the temperature below 1.5 K. The cold apertures

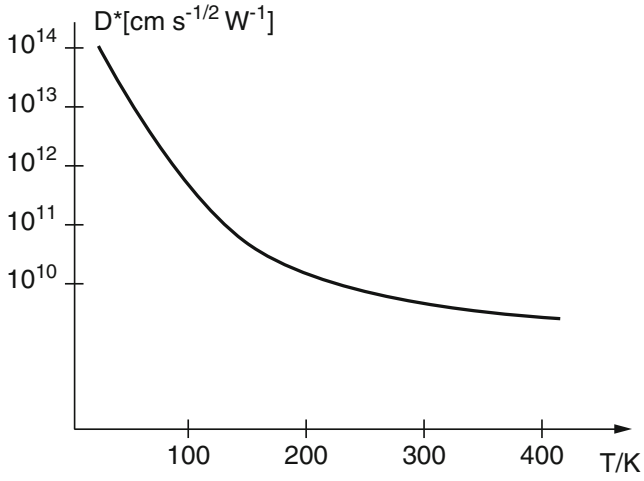

Figure 4.83 Specific detectivity D^* as a function of bolometer temperature

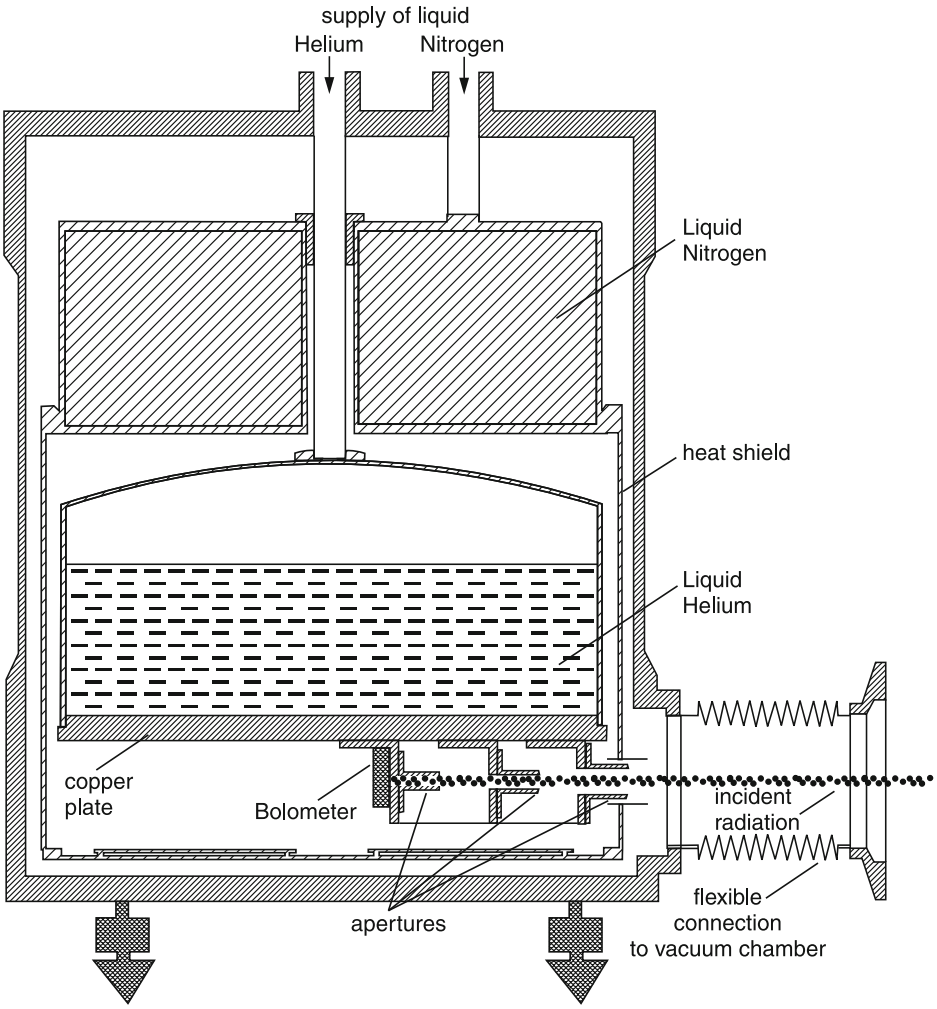

Figure 4.84 Bolometer with helium cryostat

in front of the bolometer disc stop thermal radiation from the walls of the vacuum vessel from reaching the detector. Using such a device radiation powers of less than 10^{-13} W can still be measured.

The Golay cell uses another method of thermal detection of radiation, namely the absorption of radiation in a closed gas capsule. According to the ideal gas law, the temperature rise ΔT causes the pressure rise $\Delta p = N(R/V)\Delta T$ (where N is the number of moles and R the gas constant), which expands a flexible membrane on which a mirror is mounted (Fig. 4.85a). The movement of the mirror is monitored by observing the deflection of a light beam from a light-emitting diode [227].

In modern devices the flexible membrane is part of a capacitor with the other plate fixed. The pressure rise causes a corresponding change of the capacitance, which can be converted to an AC voltage (Fig. 4.85b). This sensitive detector, which is essentially a *capacitor microphone*, is now widely used in photoacoustic spectroscopy (Vol. 2, Sect. 6.3) to detect the absorption spectrum of molecular gases by the pressure rise proportional to the absorption coefficient.

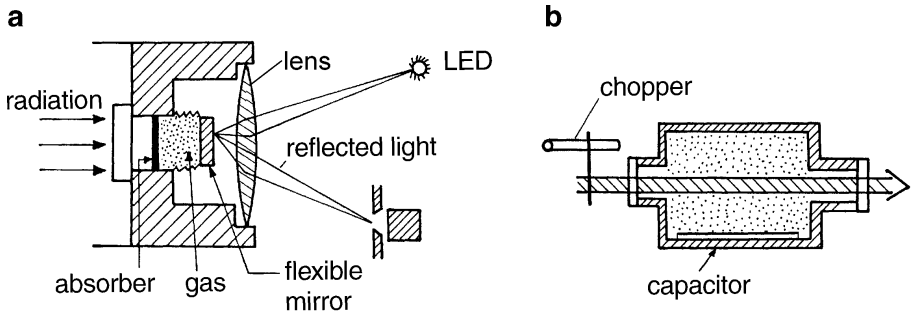

Figure 4.85 Golay cell: **a** using deflection of light by a flexible mirror; **b** monitoring the capacitance change ΔC of a capacitor C with a flexible membrane (spectraphone)

A recently developed thermal detector for infrared radiation is based on the pyroelectric effect [228–231]. Pyroelectric materials are good electrical insulators that possess an internal macroscopic electricdipole moment, depending on the temperature. The crystal neutralizes the electric field of this dielectric polarization by a corresponding surface-charge distribution. A change of the internal polarization caused by a temperature rise will produce a measurable change in surface charge, which can be monitored by a pair of electrodes applied to the sample (Fig. 4.86). Because of the capacitive transfer of the change of the electric dipole moments, pyroelectric detectors monitor only *changes* of input power. Any incident cw radiation therefore has to be chopped.

While the sensitivity of good pyroelectric detectors is comparable to that of Golay cells or high-sensitivity bolometers, they are more robust and therefore less delicate to handle. They also have a much better time resolution down into the nanosecond range [229]. The development of artificial pyroelectric materials in the form of thin films made of Gallium Nitride GaN or of Cesium Nitrate $CsNO_3$ has increased sensitivity and frequency response of pyroelectric detectors.

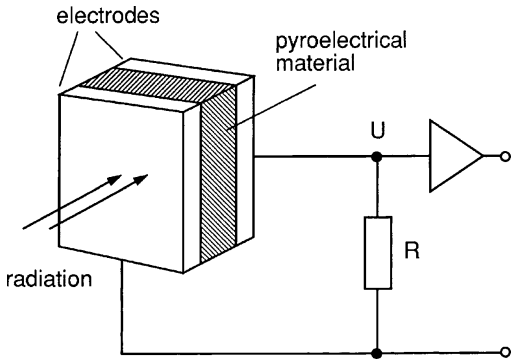

Figure 4.86 Pyroelectric detector

4.5.2 Photodiodes

Photodiodes are doped semiconductors that can be used as photovoltaic or photoconductive devices. When the p–n junction of the diode is irradiated, the photovoltage V_{ph} is generated at the open output of the diode (Fig. 4.87a); within a restricted range it is proportional to the absorbed radiation power. Diodes used as photoconductive elements change their internal resistance upon irradiation and can therefore be used as photoresistors in combination with an external voltage source (Fig. 4.87b).

For their use as radiation detectors the spectral dependence of their absorption coefficient is of fundamental importance. In an undoped semiconductor the absorption of one photon $h\nu$ causes an excitation of an electron from the valence band into the conduction band (Fig. 4.88a). With the energy gap $\Delta E_g = E_c - E_v$ between the valence and conduction band, only photons with $h\nu \geq \Delta E_g$ are absorbed. The intrinsic absorption coefficient

$$\alpha_{intr}(\nu) = \begin{cases} \alpha_0 (h\nu - \Delta E_g)^{1/2} , & \text{for} \quad h\nu > \Delta E_g , \\ 0 , & \text{for} \quad h\nu < \Delta E_g , \end{cases} \tag{4.130}$$

is shown in Fig. 4.89 for different undoped materials. The quantity α_0 depends on the material and is generally larger for semiconductors with direct transitions

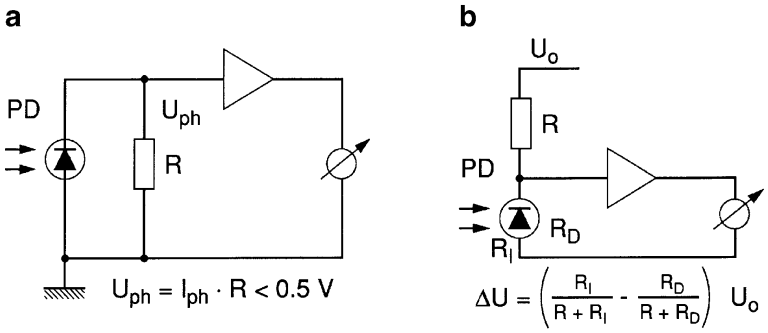

Figure 4.87 Use of a photodiode: **a** as a photovoltaic device; and **b** as a photoconductive resistor

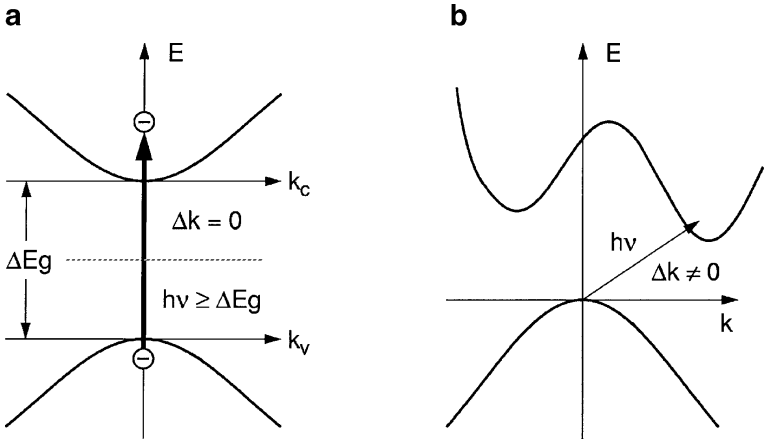

Figure 4.88 **a** Direct band–band absorption in an undoped semiconductor; and **b** indirect transitions, illustrated in a $E(k)$ band diagram

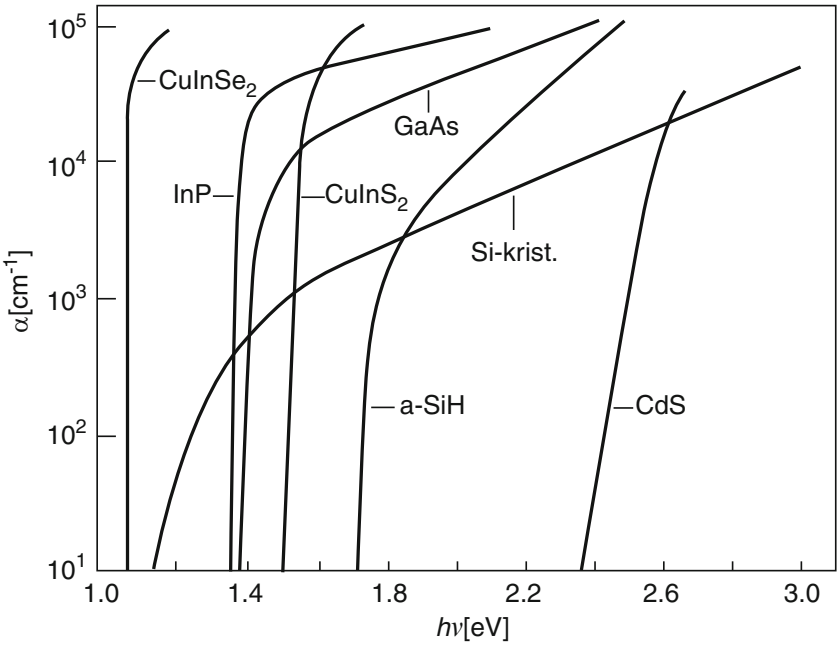

Figure 4.89 Spectral absorption $\alpha(\nu)$ of some semiconductors. Amorphous silicon a-SiH with indirect absorption transitions shows a smaller slope $d\alpha/d\nu$ of the curve $\alpha(\nu)$ while semiconductors with direct absorption (InP, GaAs or $CuInSe_2$) have a much steeper slope

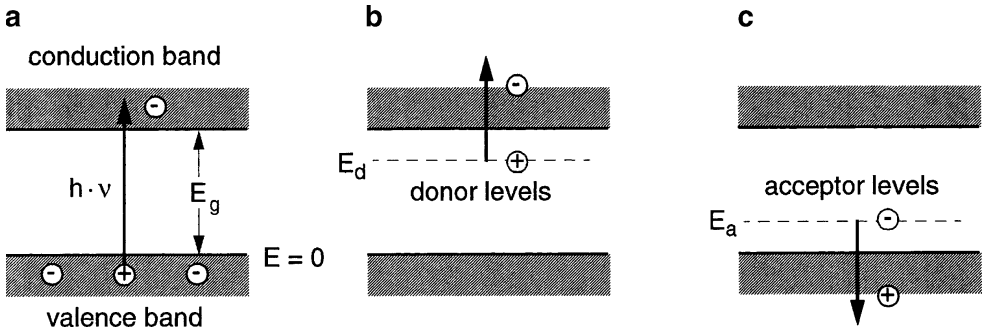

Figure 4.90 Photoabsorption in undoped semiconductors (**a**) and by donors (**b**) and acceptors (**c**) in n- or p-doped semiconductors

($\Delta k = 0$) as e.g. GaAs, than for indirect transitions with $\Delta k \neq 0$ (crystalline silicon). The steep rise of $\alpha(\nu)$ for $h\nu > E_g$ has only been observed for direct transitions, while it is much flatter for indirect transitions.

In doped semiconductors photon-induced electron transitions can occur between the donor levels and the conduction band, or between the valence band and the acceptor levels (Fig. 4.90). Since the energy gaps $\Delta E_d = E_c - E_d$ or $\Delta E_a = E_v - E_a$ are much smaller than the gap $E_c - E_v$, doped semiconductors absorb even at smaller photon energies $h\nu$ and can therefore be employed for the detection of longer wavelengths in the midinfrared. In order to minimize thermal excitation of electrons, these detectors must be operated at low temperatures. For $\lambda \leq 10\,\mu m$

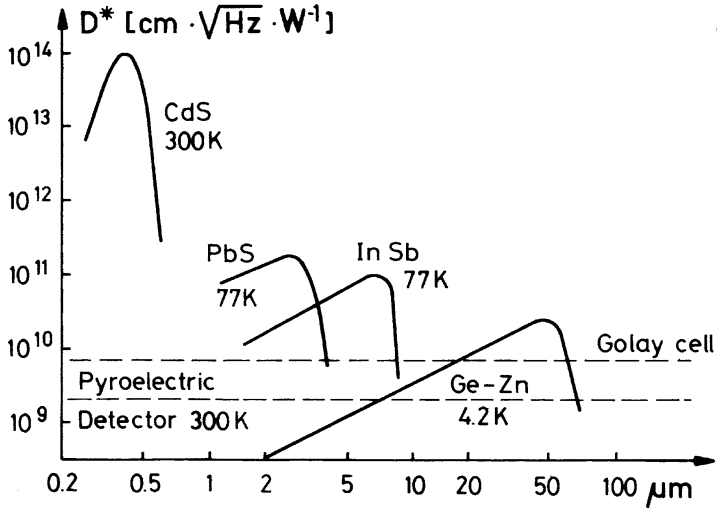

Figure 4.91 Detectivity $D^*(\lambda)$ of some photodetectors [218]

generally liquid-nitrogen cooling is sufficient, while for $\lambda > 10\,\mu\text{m}$ liquid-helium temperatures around 4–10 K are required.

Figure 4.91 plots the detectivity of commonly used photodetector materials with their spectral dependence, while Fig. 4.92 illustrates their useful spectral ranges and their dependence on the energy gap ΔE_g.

a) Photoconductive Diodes

When a photodiode is illuminated, its electrical resistance decreases from a "dark value" R_D to a value R_I under illumination. In the circuit shown in Fig. 4.87b, the change of the output voltage is given by

$$\Delta U = \left(\frac{R_D}{R_D + R} - \frac{R_I}{R_I + R} \right) U_0 = \frac{R(R_D - R_I)}{(R + R_D)(R + R_I)} U_0 \,, \qquad (4.131)$$

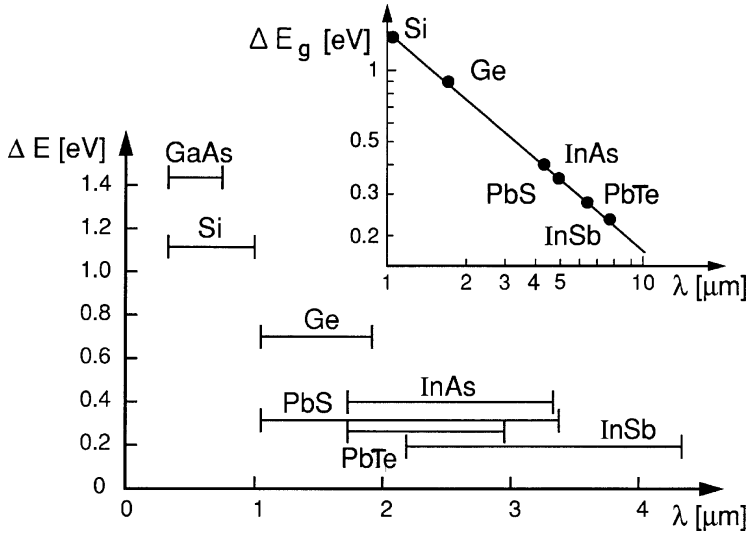

Figure 4.92 Energy gaps and useful spectral ranges of some semiconducting materials

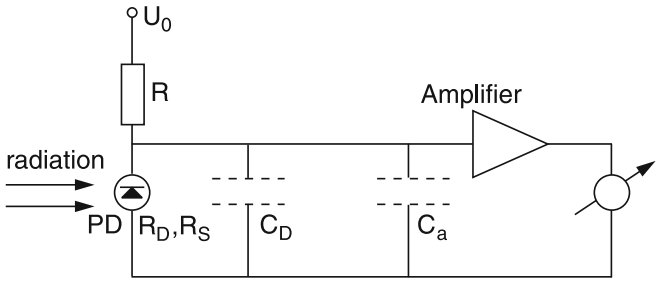

Figure 4.93 Electronic diagram of a photoconductive detector with amplifier; C_D is the capacitance of the photodiode and C_a is the capacitance of the amplifier

which becomes, at a given illumination, maximum for

$$R \approx \sqrt{R_D R_I} \, .$$

The time constant of the photoconductive diode is determined by $\tau \geq RC$, where $C = C_{PD} + C_a$ is the capacitance of the diode plus the input capacitance of the circuit. Its lower limit is set by the diffusion time of the electrons on their way from the p–n junction where they are generated to the electrodes. Detectors from PbS, for example, have typical time constants of 0.1–1 ms, while InSb detectors are much faster ($\tau \simeq 10^{-7}$–10^{-6} s). Although photoconductive detectors are generally more sensitive, photovoltaic detectors are better suited for the detection of fast signals.

b) Photovoltaic Detector

While photoconductors are passive elements that need an external power supply, photovoltaic diodes are active elements that generate their own photovoltage upon illumination, although they are often used with an external bias voltage. The principle of the photogenerated voltage is shown in Fig. 4.94.

In the nonilluminated diode, the diffusion of electrons from the n-region into the p-region (and the opposite diffusion of the holes) generates a space charge, with opposite signs on both sides of the p–n junction, which results in the diffusion voltage V_D and a corresponding electric field across the p–n junction (Fig. 4.94b). Note that this diffusion voltage cannot be detected across the electrodes of the diode, because it is just compensated by the different contact potentials between the two ends of the diode and the connecting leads.

When the detector is illuminated, electron–hole pairs are created by photon absorption within the p–n junction. The electrons are driven by the diffusion voltage into the n-region, the holes into the p-region. This leads to a *decrease* ΔV_D of the diffusion voltage, which appears as the photovoltage $V_{ph} = \Delta V_D$ across the open electrodes of the photodiode. If these electrodes are connected through an Ampéremeter, the photoinduced current

$$i_{ph} = -\eta e \phi A \, , \tag{4.132}$$

is measured, which equals the product of quantum efficiency η, the illuminated active area A of the photoiode, and the incident photon flux density $\phi = I/h\nu$.

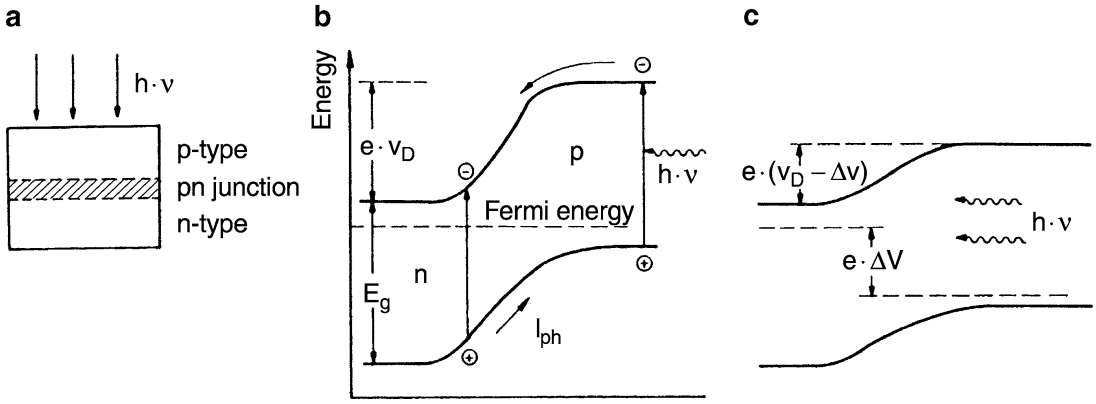

Figure 4.94 Photovoltaic diode: **a** schematic structure and **b** diffusion voltage and generation of an electron–hole pair by photon absorption within the p–n junction. **c** Reduction of the diffusion voltage V_D under illumination for an open circuit

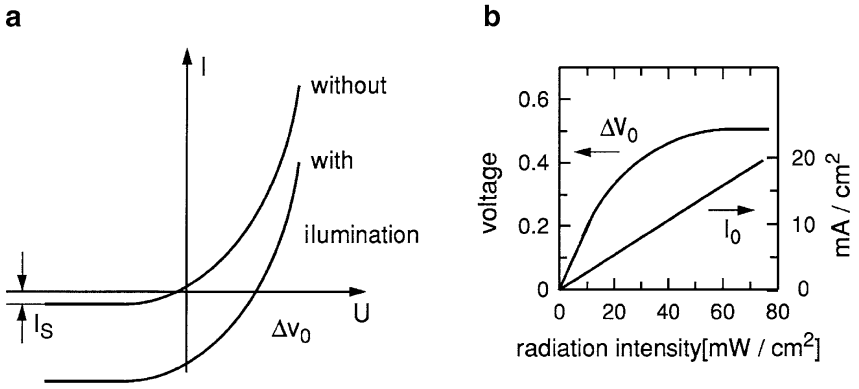

Figure 4.95 **a** Current–voltage characteristics of a dark and an illuminated diode; **b** diffusion voltage and photovoltage at the open ends and photocurrent in a shortened diode as a function of incident radiation power

The illuminated p–n photodetector can therefore be used either as a current generator or a voltage source, depending on the external resistor between the electrodes.

Note The photon-induced voltage $U_{ph} < \Delta E_g/e$ is always limited by the energy gap ΔE_g. The voltage U_{ph} across the open ends of the photodiode is reached even at relatively small photon fluxes, while the photocurrent is linear over a large range (Fig. 4.95b). When using photovoltaic detectors for measuring radiation power, the load resistor R_L must be sufficiently low to keep the output voltage $U_{ph} = i_{ph} R_L < U_s = \Delta E_g/e$ always below its saturation value U_s. Otherwise, the output signal is no longer proportional to the input power.

If an external voltage U is applied to the diode, the diode current without illumination

$$i_D(U) = CT^2 e^{-eV_D/kT}(e^{eU/kT} - 1)\ , \qquad (4.133a)$$

shows the typical diode characteristics (Fig. 4.95a). For large negative voltages U ($\exp(Ue/kT) \ll 1$), a negative reverse dark current

$$i_s = -CT^2 e^{-eV_D/kT} \tag{4.133b}$$

is flowing through the diode. During illumination the dark current i_D is superimposed by the opposite photocurrent

$$i_{ill}(U) = i_D(U) - i_{ph} \ . \tag{4.134}$$

With open ends of the diode we obtain $i = 0$, and therefore from (4.133a), (4.133b) the photovoltage becomes

$$U_{ph}(i = 0) = \frac{kT}{e}\left[\ln\left(\frac{i_{ph}}{i_s}\right) + 1\right] \ . \tag{4.135}$$

Fast photodiodes are always operated at a reverse bias voltage $U < 0$, where the saturated reverse current i_s of the dark diode is small (Fig. 4.95a). From (4.133a), (4.133b) we obtain, with $[\exp(eU/kT) \ll 1]$ for the total diode current,

$$i = -i_s - i_{ph} = -CT^2 e^{-eV_D/kT} - i_{ph} \ , \tag{4.136}$$

which becomes independent of the external voltage U.

Materials used for photovoltaic detectors are, e.g., silicon, cadmium sulfide (CdS), and gallium arsenide (GaAs). Silicon detectors deliver photovoltages up to 550 mV and photocurrents up to 40 mA/cm^2 [206]. The efficiency $\eta = P_{el}/P_{ph}$ of energy conversion reaches 10–14 %. New devices with a minimum number of crystal defects can even reach 20–30 %. Gallium arsenide (GaAs) detectors show larger photovoltages up to 1 V, but slightly lower photocurrents of about 20 mA/cm^2.

c) Fast Photodiodes

The photocurrent generates a signal voltage $V_s = U_{ph} = R_L i_{ph}$ across the load resistor R_L that is proportional to the absorbed radiation power over a large intensity range of several decades, as long as $V_s < \Delta E_g/e$ (Fig. 4.95b). From the circuit diagram in Fig. 4.96 with the capacitance C_s of the semiconductor and its series and parallel resistances R_s and R_p, one obtains for the upper frequency limit [232]

$$f_{max} = \frac{1}{2\pi C_s(R_s + R_L)(1 + R_s/R_p)} \ , \tag{4.137}$$

which reduces, for diodes with large R_p and small R_s, to

$$\boxed{f_{max} = \frac{1}{2\pi C_s R_L} \ .} \tag{4.138}$$

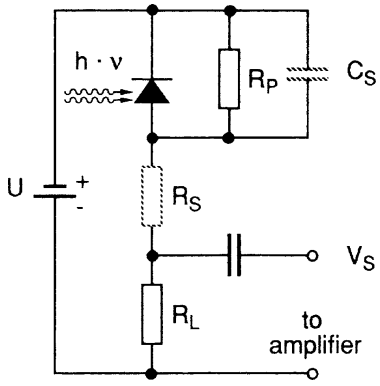

Figure 4.96 Equivalent circuit of a photodiode with internal capacity C_S, series internal resistor R_S, parallel internal resistor R_P, and external load resistor R_L

With small values of the resistor R_L, a high-frequency response can be achieved, which is limited only by the drift time of the carriers through the boundary layer of the p–n junction. This drift time can be reduced by an external bias voltage. Using diodes with large bias voltages and a 50-Ω load resistor matched to the connecting cable, rise times in the subnanosecond range can be obtained.

Example 4.25

$C_s = 10^{-11}\,\mathrm{F}, \quad R_L = 50\,\Omega \quad \Rightarrow \quad f_{max} = 300\,\mathrm{MHz}, \quad \tau = \dfrac{1}{2\pi f_{max}} \simeq 0.6\,\mathrm{ns}.$

For photon energies $h\nu$ close to the band gap, the absorption coefficient decreases, see (4.130). The penetration depth of the radiation, and with it the volume from which carriers have to be collected, becomes large. This increases the collection time and makes the diode slow.

Definite collection volumes can be achieved in *PIN diodes*, where an undoped zone I of an intrinsic semiconductor separates the p- and n-regions (Fig. 4.97). Since no space charges exist in the intrinsic zone, the bias voltage applied to the

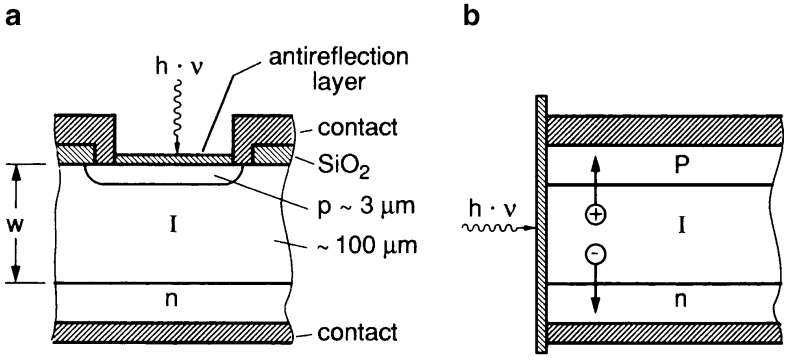

Figure 4.97 PIN photodiode with head-on (**a**) and side-on (**b**) illumination

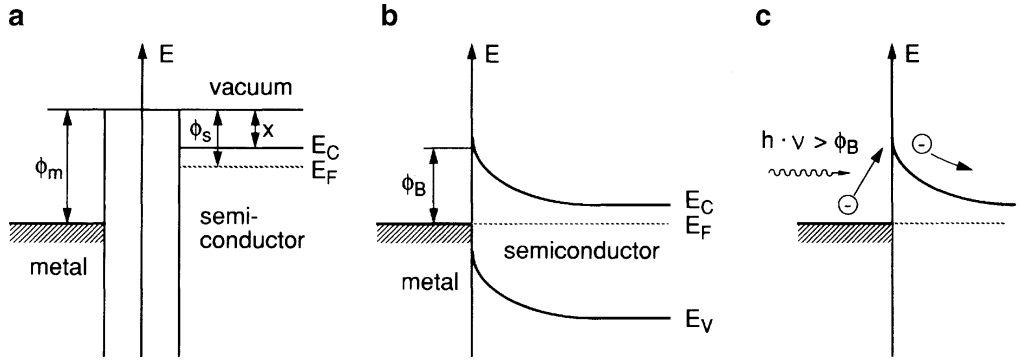

Figure 4.98 **a** Work functions ϕ_m of metal and ϕ_s of semiconductor and electron affinity χ. E_c is the energy at the bottom of the conduction band and E_F is the Fermi energy. **b** Schottky barrier at the contact layer between metal and n-type semiconductor. **c** Generation of a photocurrent

diode causes a constant electric field, which accelerates the carriers. The intrinsic region may be made quite wide, which results in a low capacitance of the p–n junction and provides the basis for a very fast and sensitive detector. The limit for the response time is, however, also set by the transit time $\tau = w/v_{th}$ of the carriers in the intrinsic region, which is determined by the width w and the thermal velocity v_{th} of the carriers. Silicon PIN diodes with a 700-µm wide zone I have response times of about 10 ns and a sensitivity maximum at $\lambda = 1.06\,\mu m$, while diodes with a 10-µm wide zone I reach 100 ps with a sensitivity maximum at a shorter wavelength–around $\lambda = 0.6\,\mu m$ [233]. Fast response combined with high sensitivity can be achieved when the incident radiation is focused from the side into the zone I (Fig. 4.97b). The only experimental disadvantage is the critical alignment necessary to hit the small active area.

Very fast response times can be reached by using the photoeffect at the metal–semiconductor boundary known as the Schottky barrier [234]. Because of the different work functions ϕ_m and ϕ_s of the metal and the semiconductor, electrons can tunnel from the material with low ϕ to that with high ϕ (Fig. 4.98). This causes a space-charge layer and a potential barrier

$$V_B = \phi_B/e\,, \quad \text{with } \phi_B = \phi_m - \chi\,, \tag{4.139}$$

between metal and semiconductor. The electron affinity is given by $\chi = \phi_s - (E_c - E_F)$. If the metal absorbs photons with $h\nu > \phi_B$, the metal electrons gain sufficient energy to overcome the barrier and "fall" into the semiconductor, which thus acquires a negative photovoltage. The *majority* carriers are responsible for the photocurrent, which ensures fast response times.

For measurements of optical frequencies, ultrafast metal–insulator-metal (MIM) diodes have been developed [235], which can be operated up to 88 THz ($\lambda = 3.39\,\mu m$). In these diodes, a 25-µm diameter tungsten wire with its end electrochemically etched to a point less than 200 nm in radius serves as the point contact element, while the optically polished surface of a nickel plate with a thin oxide layer forms the base element of the diode (Fig. 4.99).

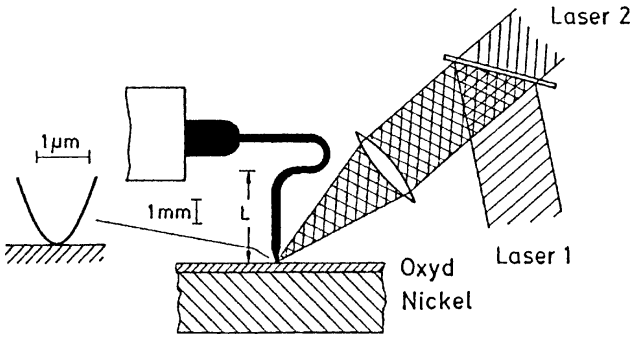

Figure 4.99 Arrangement of a metal–insulator–metal (MIM) diode used for optical frequency mixing of laser frequencies

These MIM diodes can be used as mixing elements at optical frequencies. When illuminating the contact point with a focused CO_2 laser, a response time of 10^{-14} s or better has been demonstrated by the measurement of the 88-THz emission from the third harmonic of the CO_2 laser. If the beams of two lasers with the frequencies

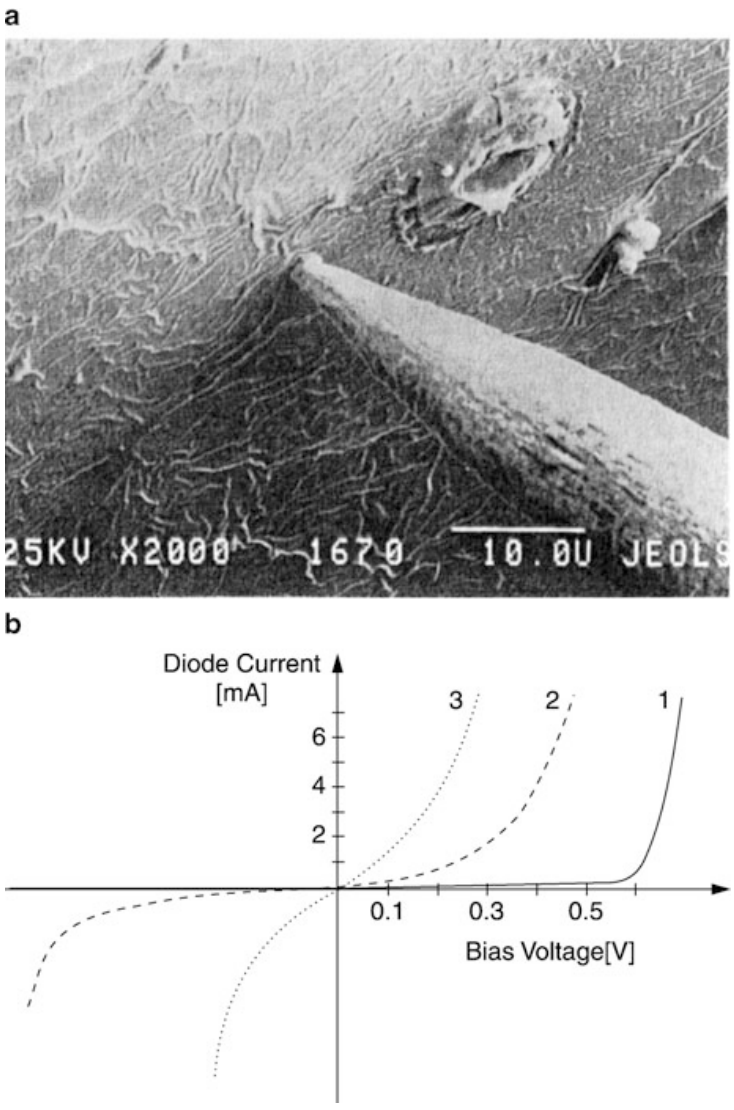

Figure 4.100 Point-contact diode: **a** electron miocroscope picture **b** current-voltage characteristics [237]

f_1 and f_2 are focused onto the junction between the nickel oxide layer and the sharp tip of the tungsten wire, the MIM diode acts as a rectifier and the wire as an antenna, and a signal with the difference frequency $f_1 - f_2$ is generated. Difference frequencies up into the terahertz range can be monitored [236] (see Sect. 6.6). The basic processes in these MIM diodes represent very interesting phenomena of solid-state physics. They could be clarified only recently [236].

Difference frequencies up to 900 GHz between two visible dye lasers have been measured with Schottky diodes (Fig. 4.100) by mixing the difference frequency with harmonics of 90-GHz microwave radiation which was also focused onto the diode [237]. Meanwhile, Schottky-barrier mixer diodes have been developed that cover the frequency range 1–10 THz [237].

d) Avalanche Diodes as Internal Amplifyers

Internal amplification of the photocurrent can be achieved with avalanche diodes, which are reverse-biased semiconductor diodes, where the free carriers acquire sufficient energy in the accelerating field to produce additional carriers on collisions with the lattice (Fig. 4.101). The multiplication factor M, defined as the average number of electron–hole pairs after avalanche multiplication initiated by a single photoproduced electron–hole pair, increases with the reverse-bias voltage. The multiplication factor

$$M = 1/\left[1 - (V/V_{\mathrm{br}})^n\right] \tag{4.140}$$

depends on the external bias voltage V and the breakdown voltage V_{br}. The value of n (2–6) depends on the material of the avalanche diode. M can be also expressed by the multiplication coefficient α for electrons and the length L of the space charge boundary:

$$M = \frac{1}{1 - \int_0^L \alpha(x)\,\mathrm{d}x} \,. \tag{4.141}$$

Values of M up to 10^6 have been reported in silicon, which allows sensitivities comparable with those of a photomultiplier. The advantage of these avalanche diodes is their fast response time, which decreases with increasing bias voltage. In this device the product of gain times bandwidth may exceed 10^{12} Hz if the breakdown voltage is sufficiently high [208]. The value of M also depends upon the temperature (Fig. 4.101b).

In order to avoid electron avalanches induced by holes accelerated into the opposite direction, which would result in additional background noise, the amplification factor for holes must be kept considerably smaller than for electrons. This is achieved by a specially tailored layer structure, which yields a sawtooth-like graded

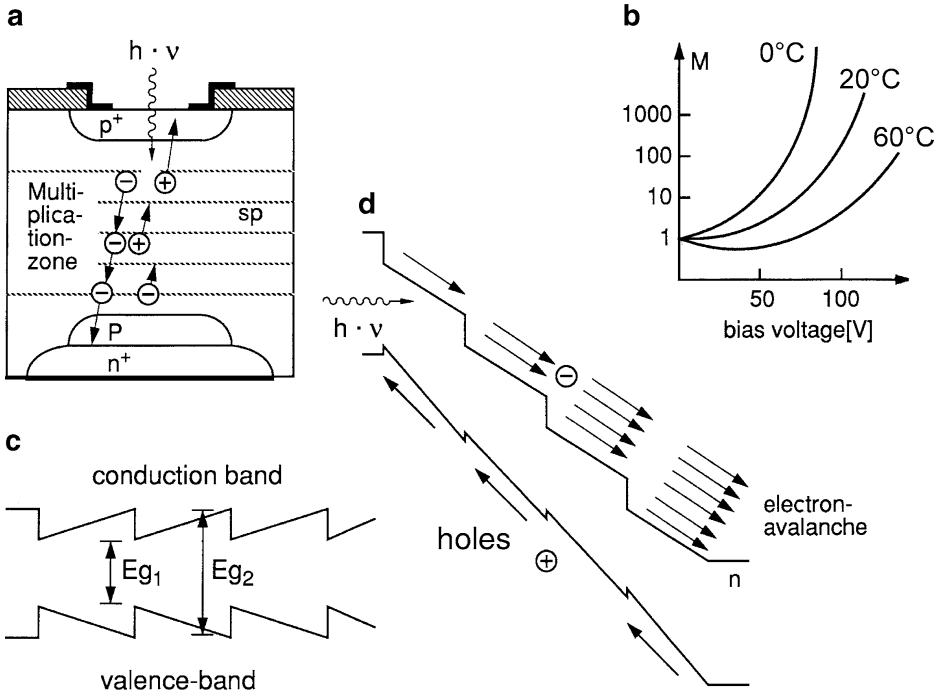

Figure 4.101 Avalanche diode: **a** schematic illustration of avalanche formation (n^+, p^+ are heavily doped layers); **b** amplification factor $M(V)$ as a function of the bias voltage V for a Si-avalanche diode; **c** spatial variation of band edges and bandgap without external field; and **d** within an external electric field

band-gap dependence $\Delta E_g(x)$ in the field x-direction (Fig. 4.101c,d). In an external field this structure results in an amplification factor M that is 50–100 times larger for electrons than for holes [238].

Such modern avalanche diodes may be regarded as the solid-state analog to photomultipliers (Sect. 4.5.5). Their advantages are a high quantum efficiency (up to 40 %) and a low supply voltage (10–100 V). Their disadvantage for fluorescence detection is the small active area compared to the much larger cathode area of photomultipliers [239–241].

Detailed data on avalanche photodiodes can be found on the homepage of Hamamatsu [242].

4.5.3 Photodiode Arrays

Many small photodiodes can be integrated on a single chip, forming a *photodiode array*. If all diodes are arranged in a line we have a onedimensional diode array, consisting of up to 2048 diodes. With a diode width $b = 15\,\mu m$ and a spacing of $d = 10\,\mu m$ between two diodes, the length L of an array of 1024 diodes becomes 25 mm with a height of about 40 μm [243].

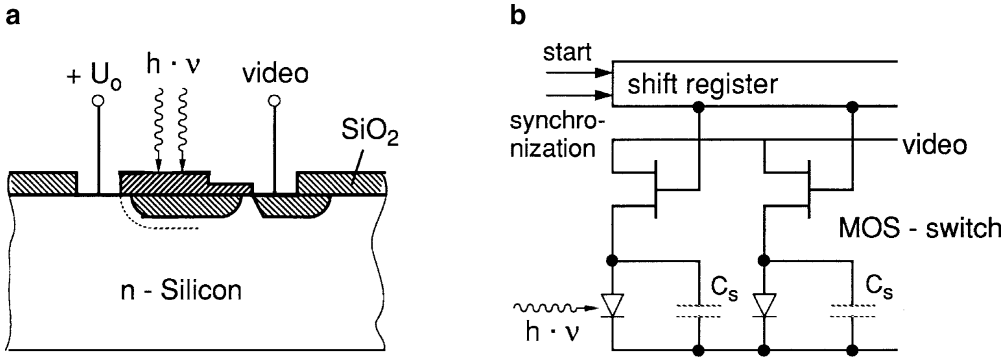

Figure 4.102 Schematic structure of a single diode within the array (**a**) and electronic circuit diagram of a one-dimensional diode array (**b**)

The basic principle and the electronic readout diagram is shown in Fig. 4.102. An external bias voltage U_0 is applied to p–n diodes with the sensitive area A and the internal capacitance C_s. Under illumination with an intensity I the photocurrent $i_{ph} = \eta A I$, which is superimposed on the dark current i_D, discharges the diode capacitance C_s during the illumination time ΔT by

$$\Delta Q = \int_t^{t+\Delta T} (i_D + \eta A I)\mathrm{d}t = C_s \Delta U \ . \tag{4.142}$$

Every photodiode is connected by a multiplexing MOS switch to a voltage line and is recharged to its original bias voltage U_0. The recharging pulse $\Delta U = \Delta Q / C_s$ is sent to a video line connected with all diodes. These pulses are, according to (4.142), a measure for the incident radiation energy $\int A I \mathrm{d}t$, if the dark current i_D is subtracted and the quantum efficiency η is known.

The maximum integration time ΔT is limited by the dark current i_D, which therefore also limits the attainable signal-to-noise ratio. At room temperature typical integration times are in the millisecond range. Cooling of the diode array by Peltier cooling down to −40 °C drastically reduces the dark current and allows integration times of 1–100 s. The minimum detectable incident radiation power is determined by the minimum voltage pulse ΔU that can be safely distinguished from noise pulses. The detection sensitivity therefore increases with decreasing temperature because of the possible increasing integration time. At room temperature typical sensitivity limits are about 500 photons per second and diode.

If such a linear diode array with N diodes and a length $L = N(b + d)$ is placed in the observation plane of a spectrograph (Fig. 4.1), the spectral interval

$$\delta\lambda = \frac{\mathrm{d}\lambda}{\mathrm{d}x} L \ ,$$

which can be detected simultaneously, depends on the linear dispersion $\mathrm{d}x/\mathrm{d}\lambda$ of the spectrograph. The smallest resolvable spectral interval

$$\Delta\lambda = \frac{\mathrm{d}\lambda}{\mathrm{d}x} b \ ,$$

is limited by the width b of the diode. Such a system of spectrograph plus diode array is called an *optical multichannel analyzer* (OMA) or an *optical spectrum analyzer* (OSA) [243, 244].

Example 4.26
$b + d = 25\,\mu\mathrm{m}, \quad L = 25\,\mathrm{mm}, \quad \mathrm{d}\lambda/\mathrm{d}x = 5\,\mathrm{nm/mm}$
$\Rightarrow \ \delta\lambda = 125\,\mathrm{nm}, \quad \Delta\lambda = 0.125\,\mathrm{nm}.$

The diodes can be also arranged in a two-dimensional array, which allows the detection of two-dimensional intensity distributions. This is, for instance, important for the observation of spatial distributions of light-emitting atoms in gas discharges or flames (Vol. 2, Sect. 15.4) or of the ring pattern behind a Fabry–Perot interferometer.

4.5.4 Charge-Coupled Devices (CCDs)

Photodiode arrays are now increasingly replaced by charge-coupled device (CCD) arrays, which consist of an array of small MOS junctions on a doped silicon substrate (Fig. 4.103) [245–250]. The incident photons generate electrons and holes in the n- or p-type silicon. The electrons or holes are collected and change the charge of the MOS capacitances. These changes of the charge can be shifted to the next MOS capacitance by applying a sequence of suitable voltage steps according to Fig. 4.103b. The charges are thus shifted from one diode to the next until they reach the last diode of a row, where they cause the voltage change ΔU, which is sent to a video line.

The quantum efficiency η of CCD arrays depends on the material used for the substrate, it reaches peak values over 90 %. The efficiency $\eta(\lambda)$ is generally larger than 20 % over the whole spectral range, which covers the region from 350–900 nm. Using fused quartz windows, even the UV and the IR from 200–1000 nm can be covered (Fig. 4.103c), and the efficiencies of most photocathodes are exceeded (Sect. 4.5.5). The spectral range of special CCDs ranges from 0.1–1000 nm. They can therefore be used in the VUV and X-ray regions, too. The highest sensitivity up to 90 % efficiency is achieved with backward-illuminated devices (Fig. 4.104). Table 4.2 compiles some relevant data for commercial CCD devices and Fig. 4.105

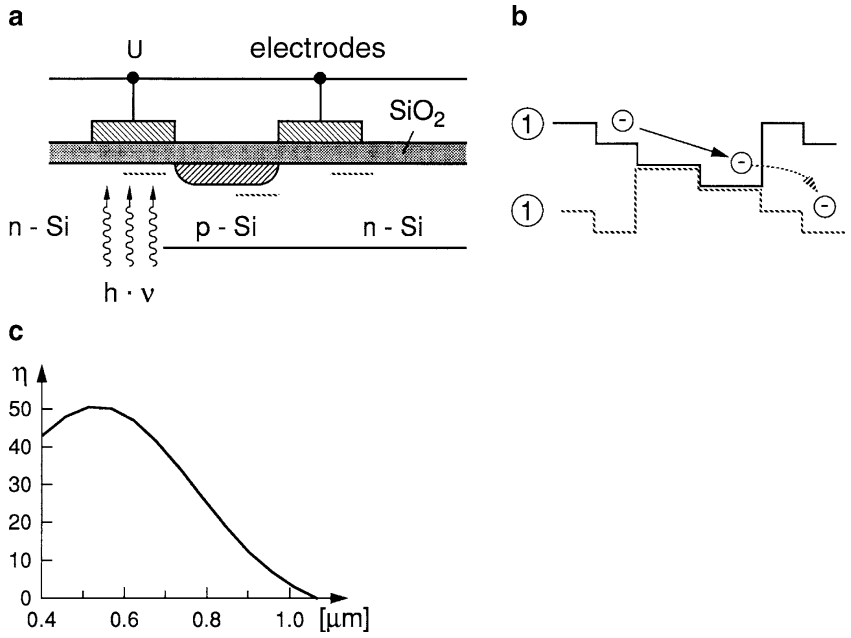

Figure 4.103 Principle of a CCD array: **a** alternately, a positive (*solid line*) and a negative (*dashed line*) voltage are applied to the electrodes. **b** This causes the charged carriers generated by photons to be shifted to the next diode. This shift occurs with the pulse frequency of the applied voltage. **c** Spectral sensitivity of CCD diodes

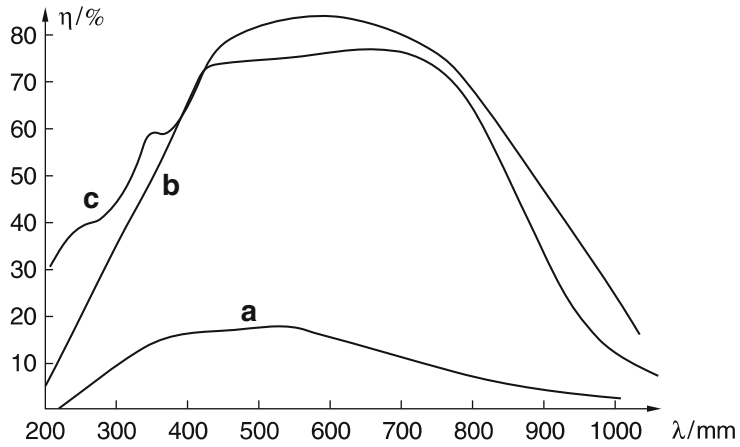

Figure 4.104 Spectral dependence of the quantum efficiency $\eta(\lambda)$ of front-illuminated (*a*) and backward-illuminated CCD arrays with visible-AR coatings (*b*) and UV–AR coatings (*c*)

compares the spectral quantum efficiency of CCD detectors with those of the photographic plate and photomultiplier cathodes.

The dark current of cooled CCD arrays may be below 10^{-2} electrons per second and diode. The readout dark pulses are smaller than those of photodiode arrays. Therefore, the sensitivity is high and may exceed that of good photomultipliers. Particular advantages are their large dynamic range, which covers about five orders of magnitude, and their linearity.

The disadvantage is their small size compared to photographic plates. This restricts the spectral range that can be detected simultaneously. More information

Table 4.2 Characteristic data of CCD arrays

Active area [mm^2]	24.6 × 24.6
Pixel size [μm]	7.5 × 15 up to 24 × 24
Number of pixels	1024 × 1024 up to 2048 × 2048
Dynamic range [bits]	16
Readout noise at 50 kHz [electron charges]	4–6
Dark charge [electrons/(h pixel)]	< 1
Hold time at −120 °C [h]	> 10
Quantum efficiency peak	
Front illuminated	50 %
Backward illuminated	> 90 %
Spectral range [mm]	300–1100

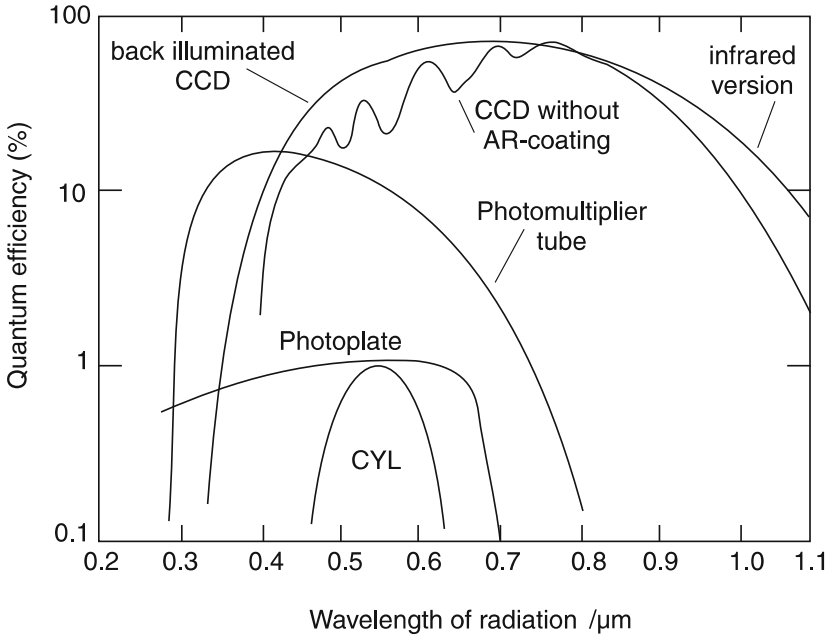

Figure 4.105 Comparison of the quantum efficiencies of CCD detectors, photoplates and photomultipliers

about CCD detectors, which are becoming increasingly important in spectroscopy, can be found in [246, 250, 251].

4.5.5 Photoemissive Detectors

Photoemissive detectors, such as the photocell or the photomuliplier, are based on the external photoeffect. The photocathode of such a detector is covered with one or several layers of materials with a low work function ϕ (e.g., alkali metal com-

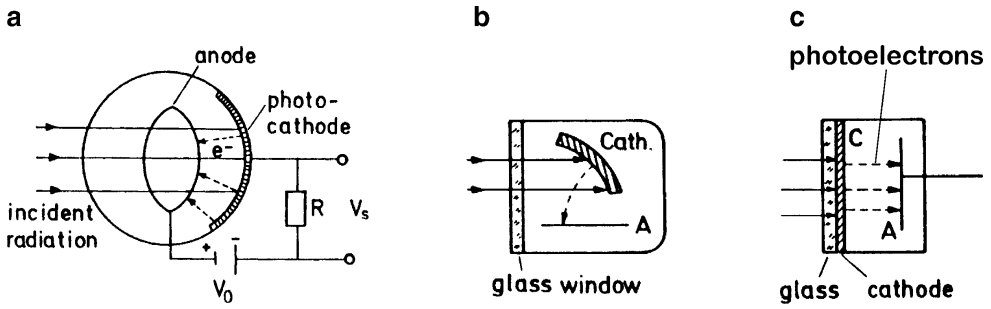

Figure 4.106 Photoemissive detector: **a** principle arrangement of a photocell; **b** opaque photocathode; and **c** semitransparent photocathode

pounds or semiconductor compounds). Under illumination with monochromatic light of wavelength $\lambda = c/\nu$, the emitted photoelectrons leave the photocathode with a kinetic energy given by the Einstein relation

$$E_{\text{kin}} = h\nu - \phi \ . \tag{4.143}$$

They are further accelerated by the voltage V_0 between the anode and cathode and are collected at the anode. The resultant photocurrent is measured either directly or by the voltage drop across a resistor (Fig. 4.106a).

a) Photocathodes

The most commonly used photocathodes are metallic or alkaline (alkali halides, alkali antimonide or alkali telluride) cathodes. The quantum efficiency $\eta = n_{\text{e}}/n_{\text{ph}}$ is defined as the ratio of the rate of photoelectrons n_{e} to the rate of incident photons n_{ph}. It depends on the cathode material, on the form and thickness of the photoemissive layer, and on the wavelength λ of the incident radiation. The quantum efficiency $\eta = n_a n_b n_c$ can be represented by the product of three factors. The first factor n_a gives the probability that an incident photon is actually absorbed. For materials with a large absorption coefficient, such as pure metals, the reflectivity R is high (e.g., for metallic surfaces $R \geq 0.8$–0.9 in the visible region), and the factor n_a cannot be larger than $(1 - R)$. For semitransparent photocathodes of thickness d, on the other hand, the absorption must be large enough to ensure that $\alpha d > 1$. The second factor n_b gives the probability that the absorbed photon really produces a photoelectron instead of heating the cathode material. Finally, the third factor n_c stands for the probability that this photoelectron reaches the surface and is emitted instead of being backscattered into the interior of the cathode.

Two types of photoelectron emitters are manufactured: opaque layers, where light is incident on the same side of the photocathode from which the photoelectrons are emitted (Fig. 4.106b); and semitransparent layers (Fig. 4.106c), where light enters at the opposite side to the photoelectron emission and is absorbed throughout the thickness d of the layer. Because of the two factors n_a and n_c, the quantum effi-

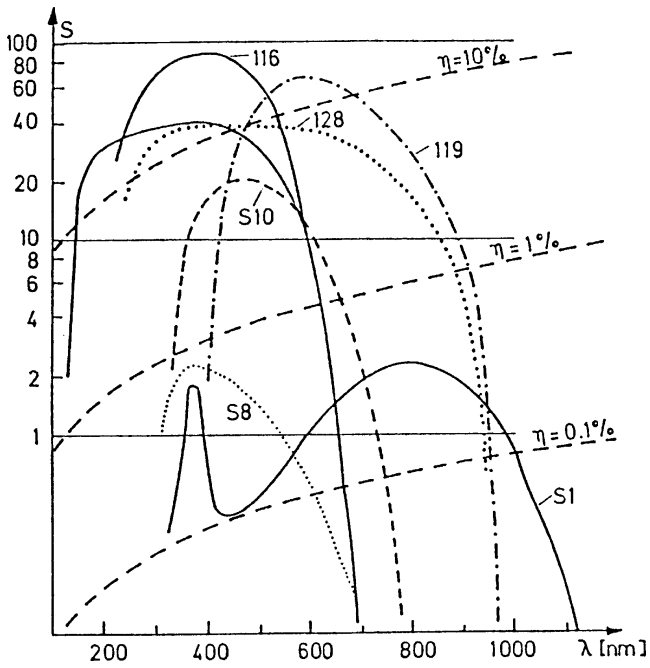

Figure 4.107 Spectral sensitivity curves of some commercial cathode types. The *solid lines* give $S(\lambda)$ [mA/W], whereas the *dashed curves* given quantum efficiencies $\eta = n_e/n_{ph}$

ciency of semitransparent cathodes and its spectral change are critically dependent on the thickness d, and reach that of the reflection-mode cathode only if the value of d is optimized.

Figure 4.107 shows the spectral sensitivity $S(\lambda)$ of some typical photocathodes, scaled in milliamperes of photocurrent per watt incident radiation. For comparison, the quantum efficiency curves for $\eta = 0.001$, 0.01 and 0.1 are also drawn (dashed curves). Both quantities are related by

$$S = \frac{i}{P_{in}} = \frac{n_e e}{n_{ph} h\nu} \Rightarrow S = \frac{\eta e \lambda}{hc} . \tag{4.144}$$

For most emitters the threshold wavelength for photoemission is below 0.85 μm, corresponding to a work function $\phi \geq 1.4$ eV. An example for such a material with $\phi \sim 1.4$ eV is a surface layer of NaKSb [252]. Only some complex cathodes consisting of two or more separate layers have an extended sensitivity up to about $\lambda \leq 1.2$ μm. For instance, an InGaAs photocathode has an extended sensitivity in the infrared, reaching up to 1700 nm. The spectral response of the most commonly fabricated photocathodes is designated by a standard nomenclature, using the symbols S1 to S20. Some newly developed types are labeled by special numbers, which differ for the different manufacturers [253]. Examples are S1 = Ag − O − Cs (300–1200 nm) or S4 = Sb − Cs (300–650 nm).

Recently, a new type of photocathode has been developed that is based on photoconductive semiconductors whose surfaces have been treated to obtain a state of negative electron affinity (NEA) (Fig. 4.108). In this state an electron at the bottom of the conduction band inside the semiconductor has a higher energy than the zero energy of a free electron in vacuum [254]. When an electron is excited by absorption of a photon into such an energy level within the bulk, it may travel to

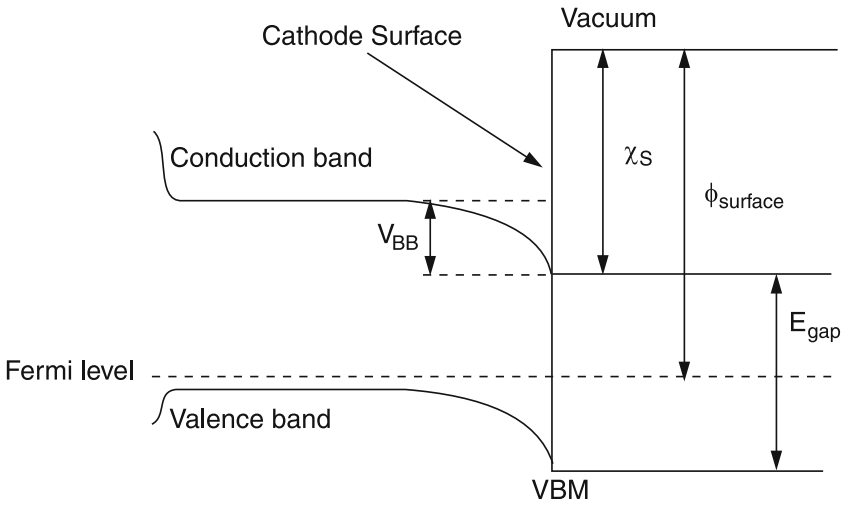

Figure 4.108 Level scheme for negative electron affinity photocathodes

the surface and leave the photocathode. These NEA cathodes have the advantage of a high sensitivity, which is fairly constant over an extended spectral range and even reaches into the infrared up to about 1.2 μm. Since these cathodes represent cold-electron emission devices, the dark current is very low. Until now, their main disadvantage has been the complicated fabrication procedure and the resulting high price.

Different devices of photoemissive detectors are of major importance in modern spectroscopy. These are the *the photomultiplier, the image intensifier, and the streak camera*.

b) Photomultipliers

Photomultipliers are a good choice for the detection of low light levels. They overcome some of the noise limitations by internal amplification of the photocurrent using secondary-electron emission from internal dynodes to mulitply the number of photoelectrons (Fig. 4.109).

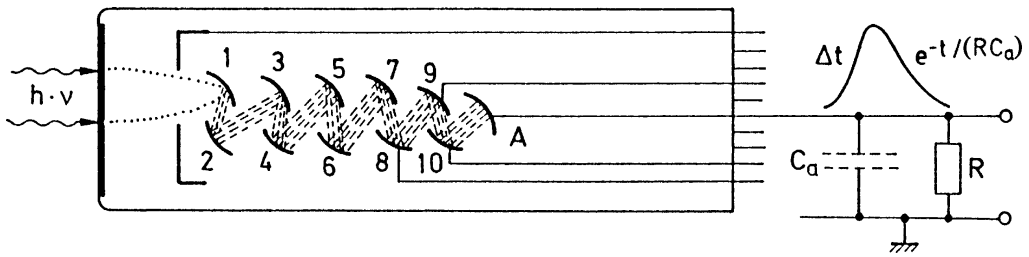

Figure 4.109 Photomultiplier with time-dependent output voltage pulse induced by an electron avalanche that was triggered by a delta-function light pulse

The photoelectrons emitted from the cathode are accelerated by a voltage of a few hundred volts and are focused onto the metal surface (e.g., Cu–Be) of the first "dynode" where each impinging electron releases, on the average, q secondary electrons. These electrons are further accelerated to a second dynode where each secondary electron again produces about q tertiary electrons, and so on. The amplification factor q depends on the accelerating voltage U, on the incidence angle α, and on the dynode material. Typical figures for $U = 200\,\mathrm{V}$ are $q = 3$–5. A photomulitplier with ten dynodes therefore has a total current amplification of $G = q^{10} \sim 10^5$–10^7. Each photoelectron in a photomultiplier with N dynodes produces a charge avalanche at the anode of $Q = q^N e$ and a corresponding voltage pulse of

$$V = \frac{Q}{C_\mathrm{a}} = \frac{q^N e}{C_\mathrm{a}} = \frac{Ge}{C_\mathrm{a}} \,, \tag{4.145}$$

where C_a is the capacitance of the anode (including connections).

Example 4.27
$G = 2 \times 10^6, \quad C_\mathrm{a} = 30\,\mathrm{pf} \;\Rightarrow\; V = 10.7\,\mathrm{mV}.$

For cw operation the DC output voltage is given by $V = i_\mathrm{a} \cdot R$, independent of the capacitance C_a.

For experiments demanding high time resolution, the rise time of this anode pulse should be as small as possible. Let us consider which effects may contribute to the anode pulse rise time, caused by the spread of transit times for the different electrons [255, 256]. Assume that a single photoelectron is emitted from the photocathode, and is accelerated to the first dynode. The initial velocities of the secondary electrons vary because these electrons are released at different depths of the dynode material and their initial energies, when leaving the dynode surface, are between 0 and 5 eV. The transit time between two parallel electrodes with distance d and potential difference V is obtained from $d = \frac{1}{2}at^2$ with $a = \mathrm{e}V/(md)$, which gives

$$t = d\sqrt{\frac{2m}{\mathrm{e}V}} \,, \tag{4.146}$$

for electrons with mass m starting with zero initial energy. Electrons with the initial energy E_kin reach the next electrode earlier by the time difference

$$\Delta t_1 = \frac{d}{\mathrm{e}V}\sqrt{2mE_\mathrm{kin}} \,. \tag{4.147}$$

Example 4.28
$E_{\mathrm{kin}} = 0.5\,\mathrm{eV}, \quad d = 1\,\mathrm{cm}, \quad V = 250\,\mathrm{V} \Rightarrow \Delta t_1 = 0.1\,\mathrm{ns}.$

The electrons travel slightly different path lengths through the tube, which causes an additional time spread of

$$\Delta t_2 = \Delta d \sqrt{\frac{2m}{eV}}\,, \tag{4.148}$$

which is of the same magnitude as Δt_1. The rise time of an anode pulse started by a single photoelectron therefore decreases with increasing voltage proportional to $V^{-1/2}$. It depends on the geometry and form of the dynode structures.

When a short intense light pulse produces many photoelectrons simultaneously, the time spread is further increased by two phenomena:

- The initial velocities of the emitted photoelectrons differ, e.g., for a cesium antimonide S5 cathode between 0 and 2 eV. This spread depends on the wavelength of the incoming light [257]a.
- The time of flight between the cathode and the first dynode strongly depends on the locations of the spot on the cathode where the photoelectron is emitted. The resulting time spread may be larger than that from the other effects, but may be reduced by a focusing electrode between the cathode and the first dynode with careful optimization of its potential. Typical anode rise times of photomultipliers range from 0.5–20 ns. For specially designed tubes with optimized side-on geometry, where the curved opaque cathode is illuminated from the side of the tube, rise times of 0.4 ns have been achieved [257]b. Shorter rise times can be reached with channel plates and channeltrons [257]c.

Example 4.29
Photomultiplier type 1P28: $N = 9, q = 5.1$ at $V = 1250\,\mathrm{V} \Rightarrow G = 2.5 \times 10^6$; anode capacitance and input capacitance of the amplifier $C_{\mathrm{a}} = 15\,\mathrm{pF}$. A single photoelectron produces an anode pulse of 27 mV with a rise time of 2 ns. With a resistor $R = 10^5\,\Omega$ at the PM exit, the trailing edge of the output pulse is $C_{\mathrm{a}} = 1.5 \times 10^{-6}\,\mathrm{s}$.

For low-level light detection, the question of noise mechanisms in photomultipliers is of fundamental importance [259]. There are three main sources of noise:

- Photomultiplier dark current;
- Noise of the incoming radiation;

- Shot noise and Johnson noise caused by fluctuations of the amplification and by noise of the load resistor.

We shall discuss these contributions separately:

- When a photomultiplier is operated in complete darkness, electrons are still emitted from the cathode. This dark current is mainly due to thermionic emission and is only partly caused by cosmic rays or by radioactive decay of spurious radioactive isotopes in the multiplier material. According to Richardson's law, the thermionic emission current

$$i = C_1 T^2 \mathrm{e}^{-C_2 \phi / T} \ , \tag{4.149}$$

strongly depends on the cathode temperature T and on its work function ϕ.
If the spectral sensitvity extends into the infrared, the work function ϕ must be small, which increases the dark current. In order to decrease the dark current, the temperature T of the cathode must be reduced. For instance, cooling a cesium–antimony cathode from 20 °C to 0 °C reduces the dark current by a factor of about ten. The optimum operation temperature depends on the cathode type (because of ϕ). For S1 cathodes, e.g., those with a high infrared sensitivity and therefore a low work function ϕ, it is advantageous to cool the cathode down to liquid nitrogen temperatures. For other types with maximum sensitivity in the green, cooling below −40 °C gives no significant improvement because the thermionic part of the dark current has already dropped below other contributions, e.g., caused by high-energy β-particles from disintegration of ^{40}K nuclei in the window material. Excessive cooling can even cause undesirable effects, such as a reduction of the signal photocurrent or voltage drops across the cathode, because the electrical resistance of the cathode film increases with decreasing temperature [260].
For many spectroscopic applications only a small fraction of the cathode area is illuminated, e.g., for photomultipliers behind the exit slit of a monochromator. In such cases, the dark current can be futher reduced either by using photomulitpliers with a small effective cathode area or by placing small magnets around an extended cathode. The magnetic field defocuses electrons from the outer parts of the cathode area. These electrons cannot reach the first dynode and do not contribute to the dark current.
- The shot noise

$$\langle i_n \rangle_\mathrm{s} = \sqrt{2e \cdot i \cdot \Delta f} \tag{4.150a}$$

of the photocurrent [259] is amplified in a photomultiplier by the gain factor G. The root-mean-square (rms) noise voltage across the anode load resistor R is therefore

$$\begin{aligned} \langle V \rangle_\mathrm{s} &= GR\sqrt{2e\, i_\mathrm{c} \Delta f} \ , \quad i_\mathrm{c} : \text{ cathode current} \ , \\ &= R\sqrt{2e\, G i_\mathrm{a} \Delta f} \ , \quad i_\mathrm{a} = G i_\mathrm{c} : \text{ anode current} \ , \end{aligned} \tag{4.150b}$$

if the gain factor G is assumed to be constant. However, generally G is not constant, but shows fluctuations due to random variations of the secondary-emission coefficient q, which is a small integer. This contributes to the total noise and multiplies the rms shot noise voltage by a factor $a > 1$, which depends on the mean value of q [261]. The shot noise at the anode is then:

$$\langle V_S \rangle = aR\sqrt{2eG\, i_a \Delta f}\ . \tag{4.150c}$$

- The Johnson noise of the load resistor R at the temperature T gives an rms-noise current

$$\langle i_n \rangle_J = \sqrt{4kT\Delta f/R} \tag{4.151a}$$

and a noise voltage

$$\langle V_n \rangle_J = R\langle i_n \rangle_J\ .$$

- From (4.150a)–(4.150c) we obtain with (4.151a) for the superposition $\langle V \rangle_{St+J} = \sqrt{\langle V \rangle_S^2 + \langle V_n \rangle_J^2}$ of shot noise and Johnson noise across the anode load resistor R at room temperature, where $4kT/e \approx 0.1\,\mathrm{V}$

$$\langle V \rangle_{J+s} = \sqrt{eR\Delta f(2RGa^2 i_a + 0.1)} \quad [\mathrm{Volt}]\ . \tag{4.151}$$

For $GR\, i_a a^2 \gg 0.05\,\mathrm{V}$, the Johnson noise can be neglected. With the gain factor $G = 10^6$ and the load resistor of $R = 10^5\,\Omega$, this implies that the anode current i_a should be larger than 5×10^{-13} A. Since the anode dark current is already much larger than this limit, we see that *the Johnson noise does not contribute to the total noise of photomultipliers*.

The channel photomultiplier is a new photomultiplier design for low-level light detection. Here the photoelectrons released from the photocathode are not multiplied by a series of dynodes, but instead move from the cathode to the anode through a curved narrow semiconductive channel (Fig. 4.110). Each time a photoelectron hits the inner surface of the channel, it releases q secondary electrons, where the integer q depends on the voltage applied between the anode and the cathode. The curved geometry causes a grazing incidence of the electrons onto the surface, which enhances the secondary emission factor q. The total gain of these channel photomultipliers (CPM) can exceed $M = 10^8$ and is therefore generally higher than for PM with dynodes.

The main advantages of the CPM are its compact design, its greater dynamic range and its lower dark current (caused mainly by thermionic emission from the photocathode) which is smaller due to its reduced area. The noise caused by fluctuations in the multiplication factor is also smaller, due to the larger value of the secondary emission factor q.

A significant improvement of the signal-to-noise ratio in detection of low levels of radiation can be achieved with single-photon counting techniques, which enable spectroscopic investigations to be performed at incident radiation fluxes

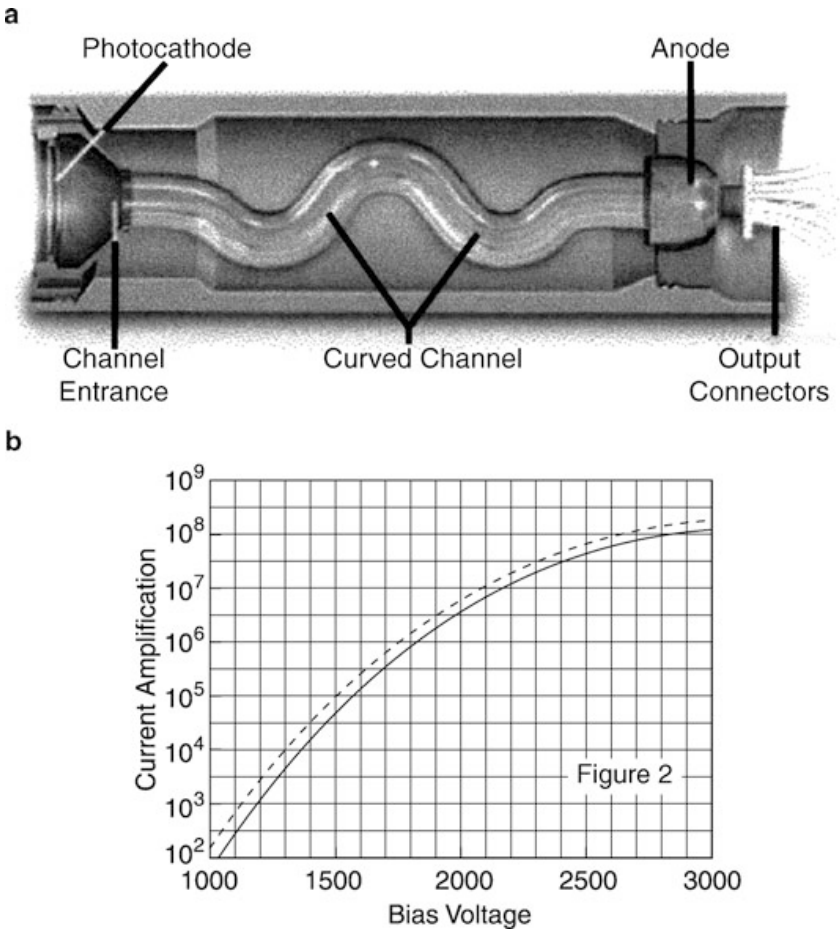

Figure 4.110 Channel photomultiplier. **a** Schematic design; **b** gain factor G as a function of the applied voltage between cathode and anode. [From Olympics Fluo View Resource Center]

down to 10^{-17} W. These techniques are discussed in Sect. 4.5.6. More details about photomultipliers and optimum conditions of performance can be found in excellent introductions issued by Hamamatsu, EMI or RCA [261, 262]. An extensive review of photoemissive detectors has been given by Zwicker [252]; see also the monographs [216, 217, 258, 263, 264].

c) Microchannel Plates

Photomultipliers are now often replaced by microchannel plates. They consist of a photocathode layer on a thin semiconductive glass plate (0.5–1.5 mm) that is perforated by millions of small holes with diameters in the range 10–25 μm (Fig. 4.111). The total area of the holes covers about 60 % of the glass plate area. The inner surface of the holes (channels) has a high secondary emission coefficient for electrons that enter the channels from the photocathode and are accelerated by a voltage applied between the two sides of the glass plate. The amplification factor

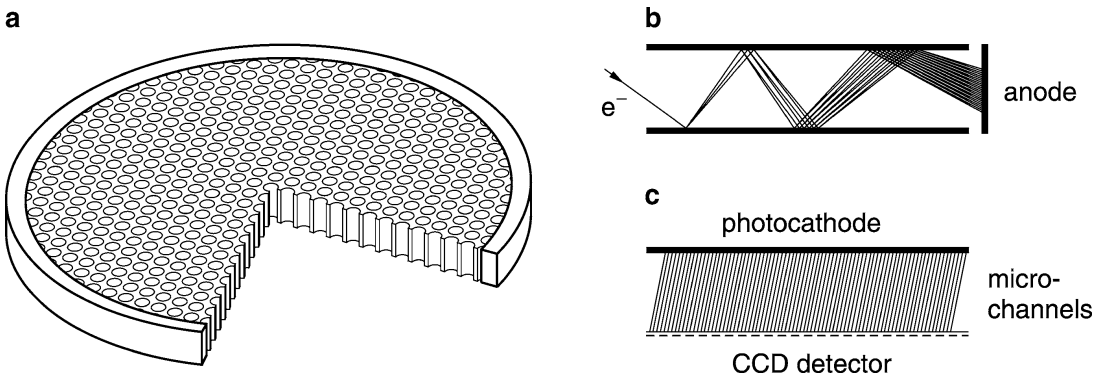

Figure 4.111 Microchannel plate (MCP): **a** schematic construction; **b** electron avalanche in one channel; **c** schematic arrangement of MCP detector with spatial resolution

is about 10^3 at an electric field of 500 V/mm. Placing two microchannel plates in series (Fig. 4.112) allows an amplification of 10^6, which is comparable to that of photomultipliers.

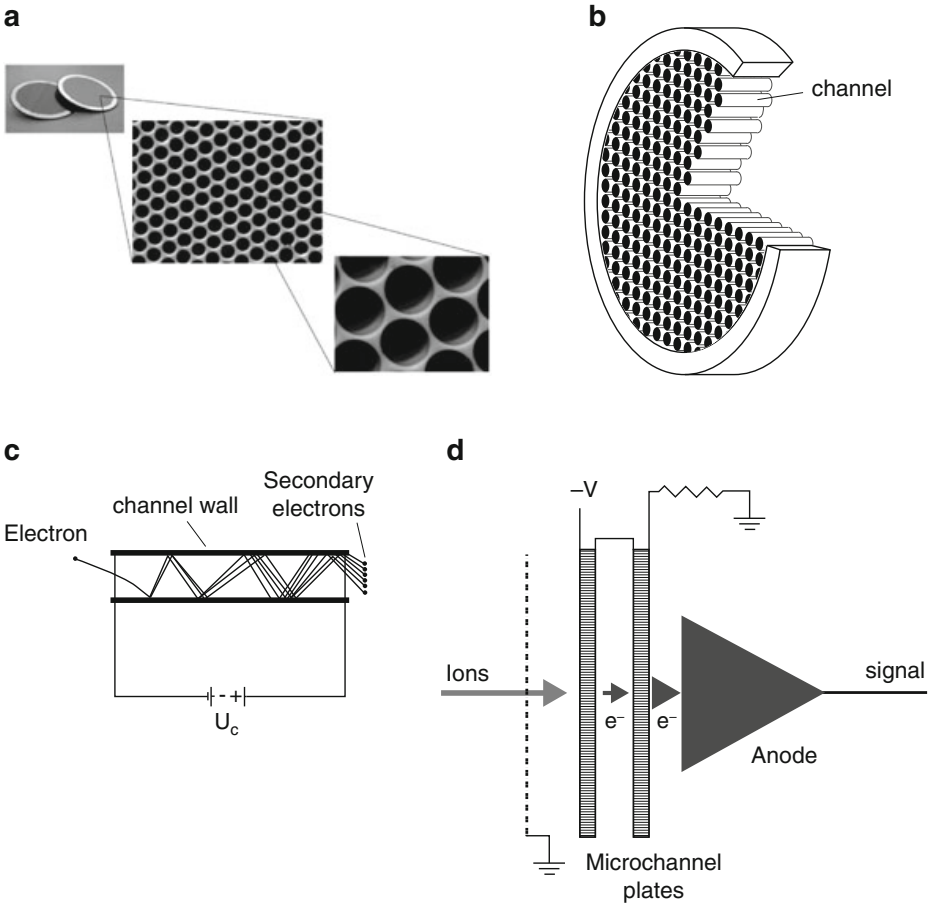

Figure 4.112 Microchannel plate: **a** microholes; **b** design of a microchannel plate; **c** principle of amplification; **d** two-stage microchannel plate

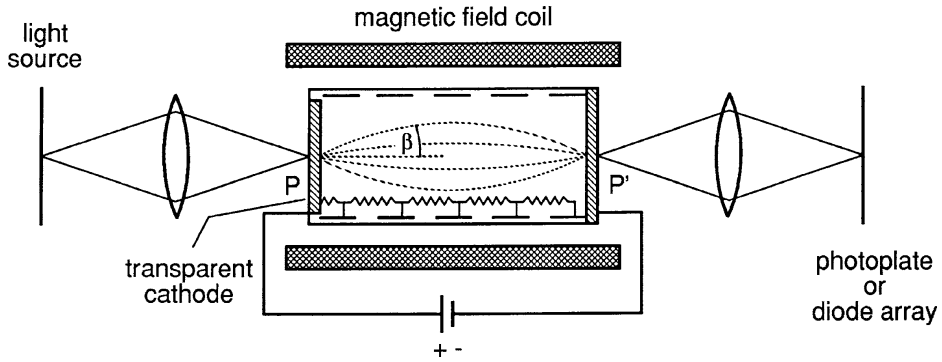

Figure 4.113 Single-stage image intensifier with magnetic focusing

The advantage of the microchannel plates is the short rise time (< 1 ns) of the electron avalanches generated by a single photon, the small size, and the possibility of spatial resolution [265].

d) Photoelectric Image Intensifiers

Image intensifiers consist of a photocathode, an electro-optical imaging device, and a fluorescence screen, where an intensified image of the irradiation pattern at the photocathode is reproduced by the accelerated photoelectrons. Either magnetic or electric fields can be used for imaging the cathode pattern onto the fluorescent screen. Instead of the intensified image being viewed on a phosphor screen, the electron image can be used in a camera tube to generate picture signals, which can be reproduced on the television screen and can be stored either photographically or on a recording medium [266–270].

For applications in spectroscopy, the following characteristic properties of image intensifiers are important:

- The intensity magnification factor M, which gives the ratio of output intensity to input intensity;
- The dark current of the system, which limits the minimum detectable input power;
- The spatial resolution of the device, which is generally given as the maximum number of parallel lines per millimeter of a pattern at the cathode which can still be resolved in the intensified output pattern;
- The time resolution of the system, which is essential for recording of fast transient input signals.

Figure 4.113 illustrates a simple, single-stage image intensifier with a magnetic field parallel to the accelerating electric field. All photoelectrons starting from the point P at the cathode follow helical paths around the magnetic field lines and are focused into P′ at the phosphor screen after a few revolutions. The location of P′ is, to a first approximation, independent of the direction β of the initial photoelec-

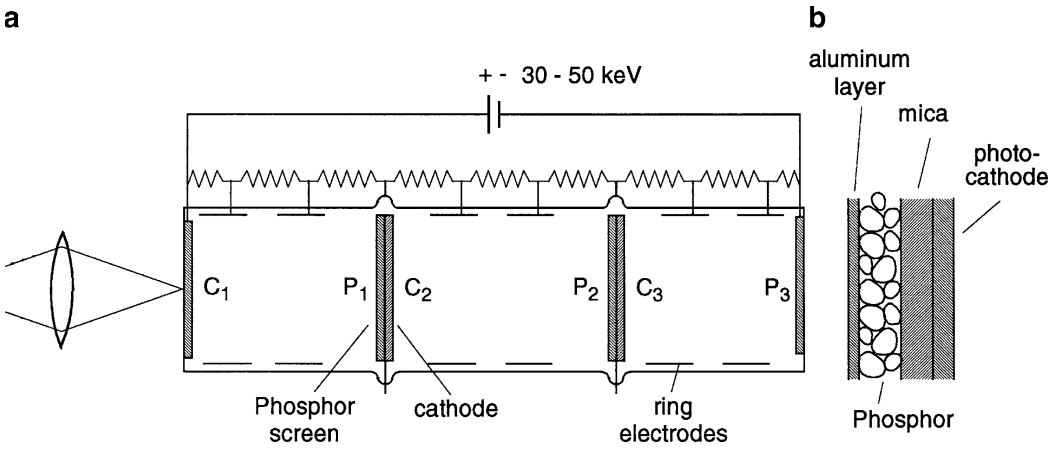

Figure 4.114 Cascade image intensifier: **a** schematic diagram with cathodes C_i, fluorescence screens P_i, and ring electrodes providing the acceleration voltage; **b** detail of phosphor–cathode sandwich structure

tron velocities. To get a rough idea about the possible magnification factor M, let us assume a quantum efficiency of 20 % for the photocathode and an accelerating potential of 10 kV. With an efficiency of 20 % for the conversion of electron energy to light energy in the phosphor screen, each electron produces about 1000 photons with $h\nu = 2\,\text{eV}$. The amplification factor M giving the number of output photons per incoming photon is then $M = 200$. However, light from the phosphor is emitted into all directions and only a small fraction of it can be collected by an optical system. This reduces the total gain factor.

The collection efficiency can be enhanced when a thin mica window is used to support the phosphor screen and photographic contact prints of the image are made. Another way is the use of fiber-optic windows.

Larger gain factors can be achieved with cascade intensifier tubes (Fig. 4.114), where two or more stages of simple image intensifiers are coupled in series [268]. The critical components of this design are the phosphor–photocathode sandwich screens, which influence the sensitivity and the spatial resolution. Since light emitted from a spot around P on the phosphor should release photelectrons from the opposite spot around P′ of the photocathode, the distance between P and P′ should be as small as possible in order to preserve the spatial resolution. Therefore, a thin layer of phosphor (a few microns) of very fine grain-size is deposited by electrophoresis on a mica sheet with a few microns thickness. An aluminum foil reflects the light from the phosphor back onto the photocathode (Fig. 4.114b) and prevents optical feedback to the preceding cathode.

The spatial resolution depends on the imaging quality, which is influenced by the thickness of the phosphor-screen–photocathode sandwiches, by the homogeneity of the magnetic field, and by the lateral velocity spread of the photoelectrons. Red-sensitive photocathodes generally have a lower spatial resolution since the initial velocities of the photoelectrons are larger. The resolution is highest at the center of

Table 4.3 Characteristic data of image intensifiers

Type	Useful diameter [mm]	Resolution [linepairs/mm]	Gain	Spectral range [nm]
RCA 4550	18	32	3×10^4	Depending on cathode type between 160 and 1000 nm
RCA C33085DP	38	40	6×10^5	
EMI 9794	48	50	2×10^5	
Hamamatsu V4435U I.I. with Multichannel plate	25	64	4×10^6	
	40	80	1×10^7	

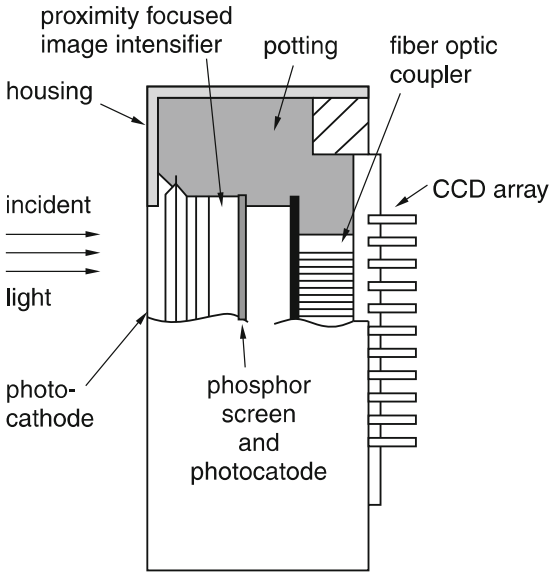

Figure 4.115 Modern version of a compact image intensifier

the screen and decreases toward the edges. Table 4.3 compiles some typical data of commercial three-stage image intensifiers [269]. In Fig. 4.115 a modern version of an image intensifier is shown. It consists of a photocathode, two short proximity-focused image intensifiers, and a fiber-optic coupler, which guides the intensified light generated at the exit of the second stage onto a CCD array.

Image intensifiers can be advantageously employed behind a spectrograph for the sensitive detection of extended spectral ranges [270]. Let us assume a linear dispersion of 1 mm/nm of a medium-sized spectrograph. An image intensifier with a useful cathode size of 30 mm and a spatial resolution of 30 lines/mm allows simultaneous detection of a spectral range of 30 nm with a spectral resolution of 3×10^{-2} nm. This sensitivity exceeds that of a photographic plate by many orders of magnitude. With cooled photocathodes, the thermal noise can be reduced to a level comparable with that of a photomultiplier, therefore incident radiation powers of a few photons can be detected. A combination of image intensifiers and vidicons or special diode arrays has been developed (optical multichannel an-

alyzers, OMA) that has proved to be very useful for fast and sensitive measurements of extended spectral ranges, in particular for low-level incident radiation (Sect. 4.5.3).

Such intensified OMA systems are commercially available. Their advantages may be summarized as follows [271, 272]:

- The vidicon targets store optical signals and allow integration over an extended period, whereas photomultipliers respond only while the radiation falls on the cathode.
- All channels of the vidicon acquire optical signals simultaneously. Mounted behind a spectrometer, the OMA can measure an extended spectral range simultaneously, while the photomultiplier accepts only the radiation passing through the exit slit, which defines the resolution. With a spatial resolution of 30 lines per mm and a linear dispersion of 0.5 nm/mm of the spectrometer, the spectral resolution is 1.7×10^{-2} nm. A vidicon target with a length of 16 mm can detect a spectral range of 8 nm simultaneously.
- The signal readout is performed electronically in digital form. This allows computers to be used for signal processing and data analyzing. The dark current of the OMA, for instance, can be automatically substracted, or the program can correct for background radiation superimposed on the signal radiation.
- Photomultipliers have an extended photocathode where the dark current from all points of the cathode area is summed up and adds to the signal. In the image intensifier in front of the vidicon, only a small spot of the photocathode is imaged onto a single diode. Thus the whole dark current from the cathode is distributed over the spectral range covered by the OMA.
  The image intensifier can be gated and allows detection of signals with high time resolution [273]. If the time dependence of a spectral distribution is to be measured, the gate pulse can be applied with variable delay and the whole system acts like a boxcar integrator with additional spectral display. The two-dimensional diode arrays also allow the time dependence of single pulses and their spectral distribution to be displayed, if the light entering the entrance slit of the spectrometer is swept (e.g., by a rotating mirror) parallel to the slit. The OMA or OSA systems therefore combine the advantages of high sensitivity, simultaneous detection of extended spectral ranges, and the capability of time resolution. These merits have led to their increased popularity in spectroscopy [271, 272].

4.5.6 Detection Techniques and Electronic Equipment

In addition to the radiation detectors, the detection technique and the optimum choice of electronic equipment are also essential factors for the success and the accuracy of spectroscopic measurements. This subsection is devoted to some modern detection techniques and electronic devices.

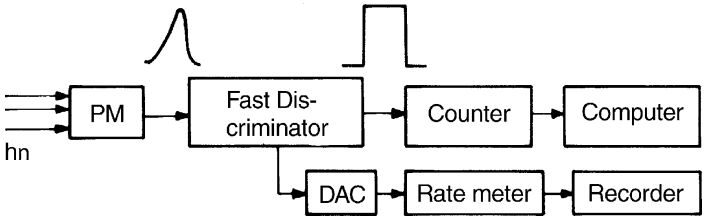

Figure 4.116 Schematic block-diagram of photon-counting electronics

a) Photon Counting

At very low incident radiation powers it is advantageous to use the photomultiplier for counting single photoelectrons emitted at a rate n per second rather than to measure the photocurrent $i = n \cdot \Delta t \cdot e \cdot G/\Delta t$ averaged over a period Δt [274]. The electron avalanches arriving at the anode with the charge $Q = Ge$ generated by a single photoelectron produce voltage pulses $U = eG/C$ at the anode with the capacitance C. With $C = 1.5 \times 10^{-11}$ F, $G = 10^6 \rightarrow U = 10$ mV. These pulses with rise times of about 1 ns trigger a fast discriminator, which delivers a TTL-norm pulse of 5 V to a counter or to a digital–analog converter (DAC) driving a rate meter with variable time constant (Fig. 4.116) [275].

Compared with the conventional analog measurement of the anode current, the photon-counting technique has the following advantages:

- Fluctuations of the photomultiplier gain G, which contribute to the noise in analog measurements, see (4.151), are not significant here, since each photoelectron induces the same normalized pulse from the discriminator as long as the anode pulse exceeds the discriminator threshold.
- Dark curent generated by thermal electrons from the various dynodes can be suppressed by setting the discriminator threshold correctly. This discrimination is particularly effective in photomultipliers with a large conversion efficiency q at the first dynode, covered with a GaAsP layer.
- Leakage currents between the leads in the photomulitplier socket contribute to the noise in current measurements, but are not counted by the discriminator if it is correctly biased.
- High-energy β-particles from the disintegration of radioactive isotopes in the window material and cosmic ray particles cause a small, but nonnegligible, rate of electron bursts from the cathode with a charge $n \cdot e$ of each burst ($n \gg 1$). The resulting large anode pulses cause additional noise of the anode current. They can, however, be completely suppressed by a window discriminator used in photon counting.
- The digital form of the signal facilitates its further processing. The discriminator pulses can be directly fed into a computer that analyzes the data and may control the experiment [276].

The upper limit of the counting rate depends on the time resolution of the discriminator, which may be below 10 ns. This allows counting of randomly distributed pulse rates up to about 10 MHz without essential counting errors.

The lower limit is set by the dark pulse rate [277]. With selected low-noise photomultipliers and cooled cathodes, the dark pulse rate may be below 1 per second. Assuming a quantum efficiency of $\eta = 0.2$, it should therefore be possible to achieve, within a measuring time of 1 s, a signalto-noise ratio of unity even at a photon flux of 5 photons/s. At these low photon fluxes, the probability $p(N)$ of N photoelectrons being detected within the time interval Δt follows a Poisson distribution

$$p(N) = \frac{\overline{N}^N \mathrm{e}^{-\overline{N}}}{N!} , \tag{4.152}$$

where $\overline{N}$ is the average number of photoelectrons detected within a given time interval Δt [277]. If the probability that at least one photoelectron will be detected within Δt is 0.99, then $1 - p(0) = 0.99$ and

$$p(0) = \mathrm{e}^{-\bar{N}} = 0.01 , \tag{4.153}$$

which yields $\overline{N} \geq 4.6$. This means that we can expect a pulse during the observation time with 99 % certainty only if at least 20 photons fall onto the photocathode with a quantum efficiency of $\eta = 0.2$. For longer detection times, however, the detectable photoelectron rate may be even lower than the dark current rate if, for instance, lock-in detection is used. It is not the dark pulse rate N_D itself that limits the signal-to-noise ratio, but rather its fluctuations, which are proportional to $N_\mathrm{D}^{1/2}$.

Because of their low noise, channel photomultipliers or avalanche diodes are well suited to low-level photon counting.

b) Measurements of Fast Transient Events

Many spectroscopic investigations require the observation of fast transient events. Examples are lifetime measurements of excited atomic or molecular states, investigations of collisional relaxation, and studies of fast laser pulses (Vol. 2, Chap. 6). Another example is the transient response of molecules when the incident light frequency is switched into resonance with molecular eigenfrequencies (Vol. 2, Chap. 7). Several techniques are used to observe and to analyze such events and recently developed instruments help to optimize the measuring procedure. The combination of a CCD detector and a gated microchannel plate, which acts as an image intensifier with nanosecond resolution, allows the time-resolved sensitive detection of fast events. In addition, there are several devices that are particularly suited for the electronic handling of short pulses. We briefly present three examples of such equipment: the *boxcar integrator* with signal averaging, the *transient recorder*, and the *fast transient digitizer* with subnanosecond resolution.

The *boxcar integrator* measures the amplitudes and shapes of signals with a constant repetition rate integrated over a specific sampling interval Δt. It records these

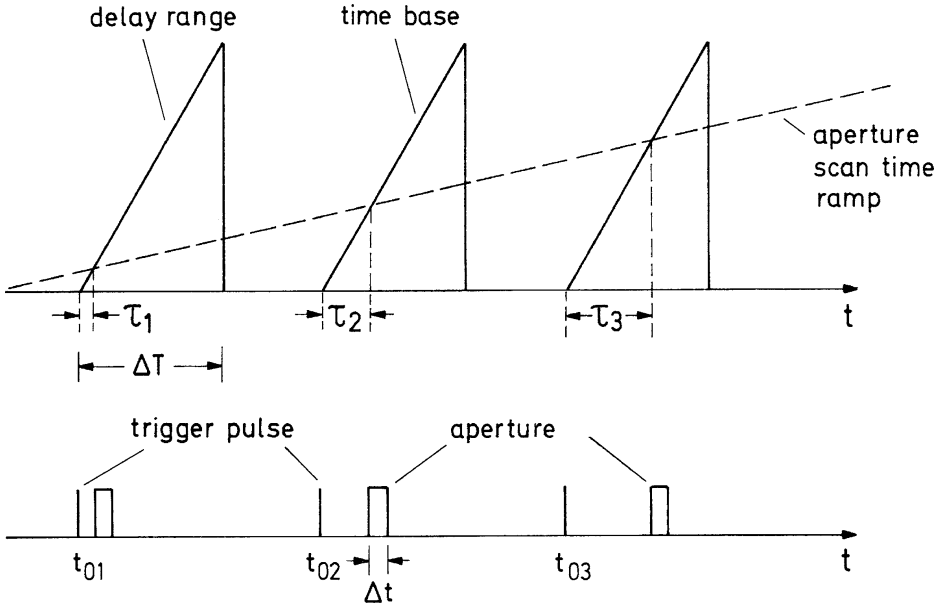

Figure 4.117 Principle of boxcar operation with synchronization of the repetitive signals. The time base determines the opening times of the gate with width Δt. The slow scan-time ramp shifts the delay times τ_i continuously over the signal–pulse time profile

signals repetitively over a selected number of pulses and computes the average value of those measurements. With a synchronized trigger signal it can be assured that one looks each time at the identical time interval of each sampled waveform. A delay circuit permits the sampled time interval Δt (called *aperture*) to be shifted to any portion of the waveform under investigation. Figure 4.117 illustrates a possible way to perform this sampling and averaging. The aperture delay is controlled by a ramp generator, which is synchronized to the signal repetition rate and which provides a sawtooth voltage at the signal repetition frequency. A slow aperture-scan ramp shifts the gating time interval Δt, where the signal is sampled at the time delay τ_i after the trigger pulse for a time interval Δt. Between two successive signals the gate time is shifted by an amount $\Delta\tau$, which depends on the slope of the ramp. This slope has to be sufficiently slow in order to permit a sufficient number of samples to be taken in each segment of the waveform. The output signal is then averaged over several scans of the time ramp by a signal averager [278]. This increases the signal-to-noise ratio and smooths the dc output, which follows the shape of the waveform under study.

The slow ramp is generally not a linearly increasing ramp as shown in Fig. 4.117, but rather a step function where the time duration of each step determines the number of samples taken at a given delay time τ. If the slow ramp is replaced by a constant selectable voltage, the system works as a gated integrator.

The integration of the input signal $U_s(t)$ over the sampling time interval Δt can be performed by charging a capacitance C through a resistor R (Fig. 4.118), which

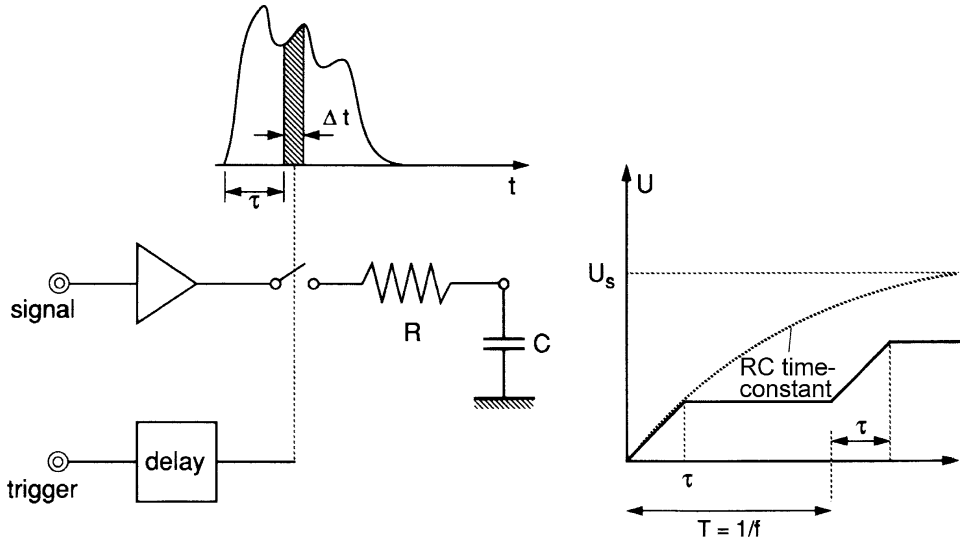

Figure 4.118 Simplified diagram of boxcar realization

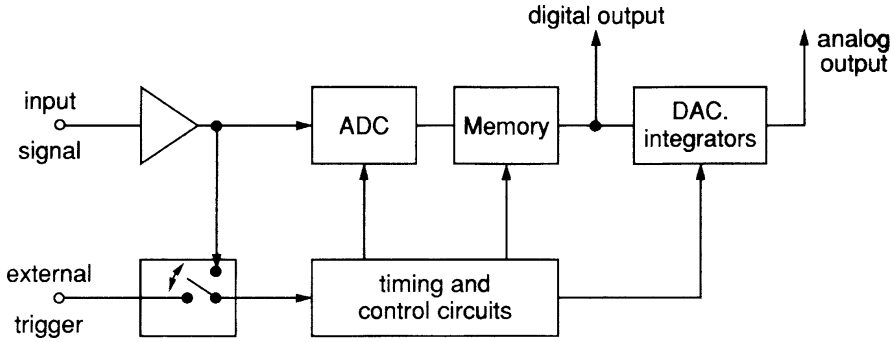

Figure 4.119 Block diagram of a transient recorder

gives a current $i(t) = U_s(t)/R$. The output is then

$$U(\tau) = \frac{1}{C} \int_{\tau}^{\tau+\Delta t} i(t)\mathrm{d}t = \frac{1}{RC} \int_{\tau}^{\tau+\Delta t} U_s(t)\mathrm{d}t \ . \tag{4.154}$$

For repetitive scans, the voltages $U(\tau)$ can be summed. Because of inevitable leakage currents, however, unwanted discharge of the capacitance occurs if the signal under study has a low duty factor and the time between successive samplings becomes large. This difficulty may be overcome by a digital output, consisting of a two-channel analog-todigital-to-analog converter. After a sampling switch opens, the acquired charge is digitized and loaded into a digital storage register. The digital register is then read by a digital-to-analog converter producing a dc voltage equal to the voltage $U(\tau) = Q(\tau)/C$ on the capacitor. This dc voltage is fed back to the integrator to maintain its output potential until the next sample is taken.

The boxcar integrator needs repetitive waveforms because it samples each time only a small time interval Δt of the input pulse and composes the whole period of

the repetitive waveform by adding many sampling points with different delays. For many spectroscopic applications, however, only single-shot signals are available. Examples are shock-tube experiments or spectroscopic studies in laser-induced fusion. In such cases, the boxcar integrator is not useful and a *transient recorder* is a better choice. This instrument uses digital techniques to sample N preselected time intervals Δt_i which cover the total time $T = N\Delta t$ of an analog signal as it varies with time. The wave shape during the selected period of time is recorded and held in the instrument's memory until the operator instructs the instrument to make a new recording. The operation of a transient recorder is illustrated in Fig. 4.119 [279, 280]. A trigger, derived from the input signal or provided externally, initiates the sweep. The amplified input signal is converted at equidistant time intervals to its digital equivalent by an analog-to-digital converter and stored in a semiconductor memory in different channels. With 100 channels, for instance, a single-shot signal is recorded by 100 equidistant sampling intervals. The time resolution depends on the sweep time and is limited by the frequency response of the transient recorder. Sample intervals between 10 ns up to 20 s can be selected. This allows sweep times of 20 µs to 5 h for 2000 sampling points. With modern devices, sampling rates of up to 500 MHz are achievable.

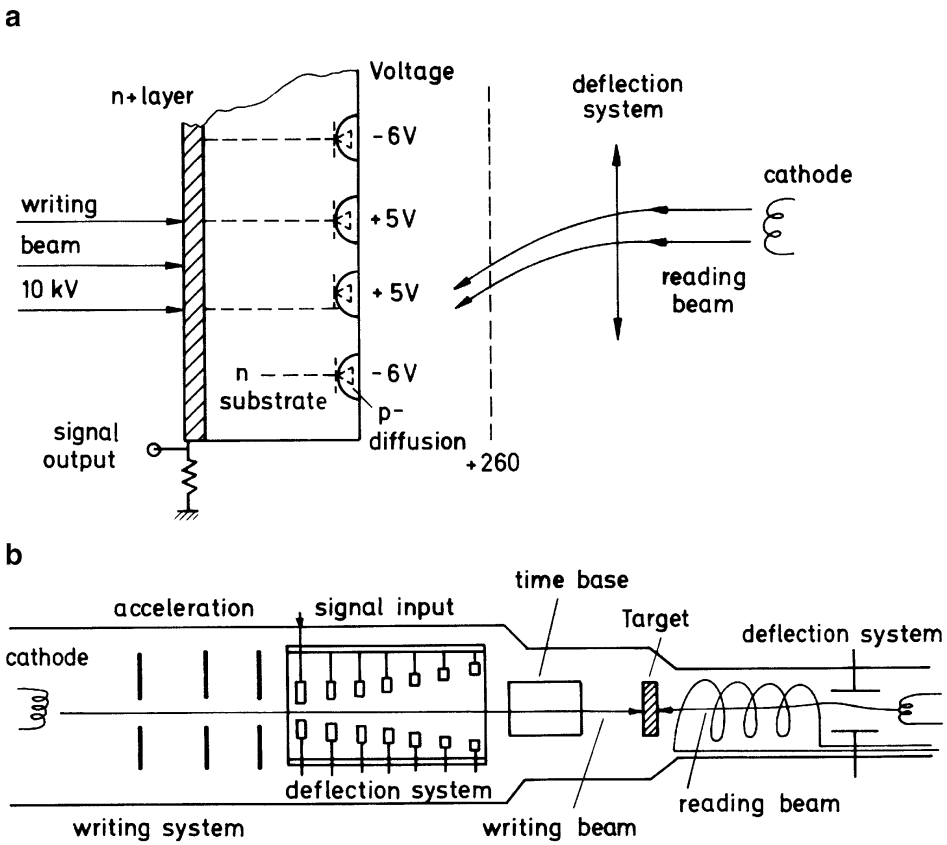

Figure 4.120 Fast transient digitizer: **a** silicon diode-array target; and **b** writing and reading gun [281]

Acquisition and analysis beyond 500 MHz has become possible by combining the features of a transient recorder with the fast response time of an electron beam that writes and stores information on a diode matrix target in a scan converter tube. Figure 4.120 illustrates the basic principle of the *transient digitizer* [281]. The diode array of about 640,000 diodes is scanned by the reading electron beam, which charges all reverse-biased p–n junctions until the diodes reach a saturation voltage. The writing electron beam impinges on the other side of the 10 µm thick target and creates electron–hole pairs, which diffuse to the anode and partially discharge it. When the reading beam hits a discharged diode, it becomes recharged, subsequently a current signal is generated at the target lead, which can be digitally processed.

The instrument can be used in a nonstoring mode where the operation is similar to that of a conventional television camera with a video signal, which can be monitored on a TV monitor. In the digital mode the target is scanned by the reading beam in discrete steps. The addresses of points on the target are transferred and stored in memory only when a trace has been written at those points on the target. This transient digitizer allows one to monitor fast transient signals with a time resolution of 100 ps and to process the data in digital form in a computer. It is, for instance, possible to obtain the frequency distribution of the studied signal from its time distribution by a Fourier transformation performed by the computer.

c) Optical Oscilloscope

The optical oscilloscope represents a combination of a streak camera and a sampling oscilloscope. Its principle of operation is illustrated by Fig. 4.121 [282]: The incident light $I(t)$ is focused onto the photocathode of the streak camera. The electrons released from the cathode pass between two deflecting electrodes toward the sampling slit. Only those electrons that traverse the deflecting electric field at a given selectable time can pass through the slit. They impinge on a phosphor screen and produce light that is detected by a photomultiplier (PM). The PM output is amplified and fed into a sampling oscilloscope, where it is stored and processed. The sampling operation can be repeated many times with different delay times t between the trigger and the sampling, similar to the principle of a boxcar operation (Fig. 4.117). Each sampling interval yields the signal

$$S(t, \Delta t) = \int_{t}^{t+\Delta t} I(t)\mathrm{d}t \ . \tag{4.155}$$

The summation over all sampled time intervals Δt gives the total signal

$$S(t) = \sum_{n=1}^{N} \int_{t=(n-1)\Delta t}^{t=n\cdot\Delta t} I(t)\mathrm{d}t \ , \tag{4.156}$$

which reflects the input time profile $I(t)$ of the incident light.

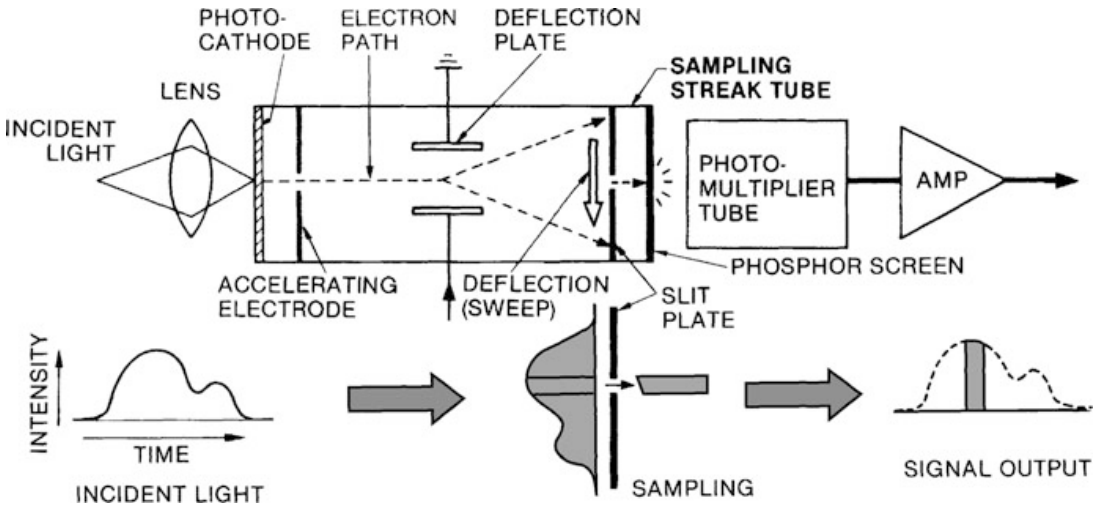

Figure 4.121 Optical oscilloscope [282]

The spectral response of the system depends on that of the first photocathode and reaches from 350 to 850 µm for the visible version and from 400 to 1550 µm for the extended infrared version. The time resolution is better than 10 ps and the sampling rate can be selected up to 2 MHz. The limitation is given by the time jitter, which was stated to be less than 20 ps.

d) Vidicon

In the Vidicon the electrons, emitted from the photocathode are accelerated by an electric field and are imaged onto a CCD array or a photodiode array. (Fig. 4.122). They produce electron–hole pairs. The voltage applied to the photodiodes drives the electrons to the positive electrode and the holes to the negative electrode resulting in a discharge of the diode capacitance. An electron beam imaged by an appropriate electron optics onto the different diodes of the array recharges the capacities up to their original voltage. This recharging current appears as voltage pulse on the common video line and gives the wanted signal.

Such photodiode arrays with the vidicon technique reach high sensitivities which are only limited by the quantum efficiency of the photocathode camparable to that of good photomultipliers. The advantage of these devices is the spatial resolution because each point on the photocathode is imaged onto a specific photodiode. If the light falling onto the photocathode has been dispersed by a spectrometer with a dispersion $\mathrm{d}\lambda/\mathrm{d}x$, the spatial resolution Δx gives a spectral resolution

$$\Delta\lambda = \frac{\mathrm{d}\lambda}{\mathrm{d}x}\,\Delta x\ .$$

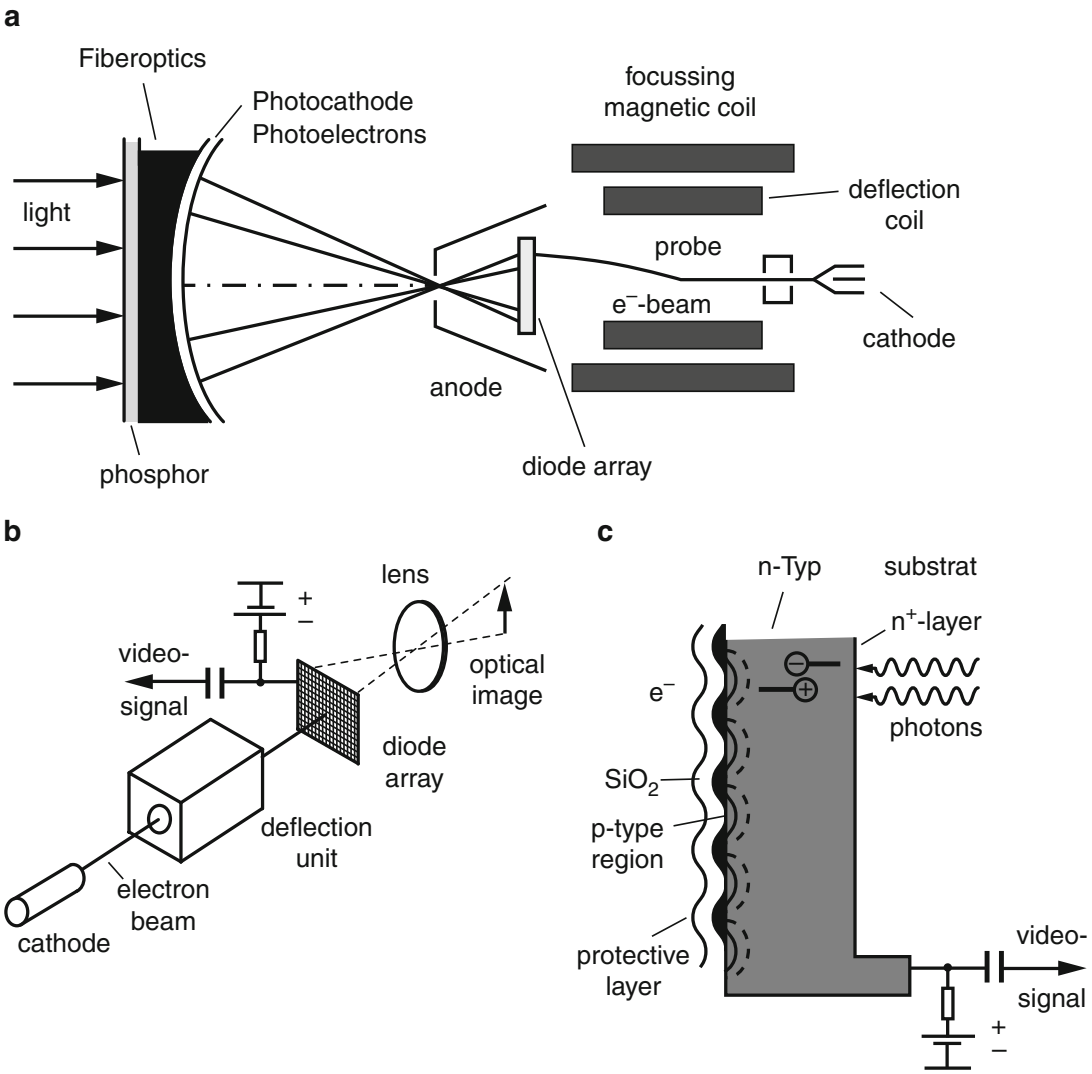

Figure 4.122 Basic principles of a vidicon. **a** Total setup, **b** details of detection system, **c** photodiode array

With a total legth L of the array the spectral range which can be simultaneously measured is

$$\delta\lambda = \frac{\mathrm{d}\lambda}{\mathrm{d}x} L \ .$$

These devices which store and analyse the information contained in the incident light simultaneously in many channels, are called *optical multichannel analysers* (OMA) or OSA (*optical spectrum analysers*). They measure two-dimensional images from weak extended light sources or cover simultaneously a spectral range $\delta\lambda$ [283]. The signals can be time integrated over a period of many seconds (for cooled devices even several hours) and are therefore superior to photomultipliers. They have found increasing importance in astronomy, where the faint images of distant galaxies are observes and their spectra are measured.

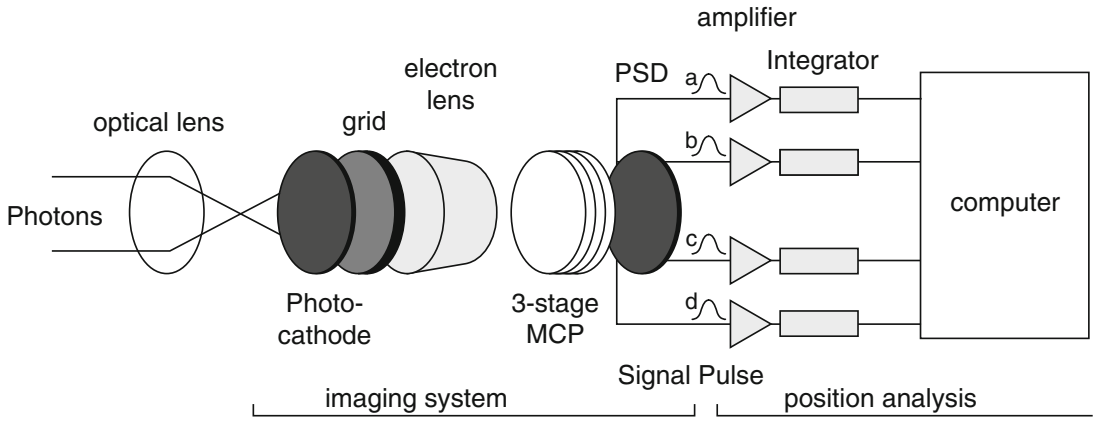

Figure 4.123 Schematic diagram of the PIAS as image intensifier with photon counting detection

e) PIAS System

PIAS is an acronym for *Photocounting Image Aquisation System*. Its basic principle is illustrated in Fig. 4.123. The photons from a light source fall onto the photocathode of an image intensifier where they release electrons which are amplified in a multichannel canal plate (MCP) and are imaged onto a position sensitive detector. A computer software analyses the data and gives the measured spectrum or the spatial variation of the intensity from extended sources [284, 285].

4.6 Conclusions

The aim of this chapter was to provide a general background in spectroscopic instrumentation, to summarize some basic ideas of spectroscopy, and to present some important relations between spectroscopic quantities. This background should be helpful in understanding the following chapters that deal with the main subject of this textbook: the applications of lasers to the solution of spectroscopic problems. Although until now we have only dealt with general spectroscopy, the examples given were selected with special emphasis on laser spectroscopy. This is especially true in Chap. 4, which is, of course, not a complete account of spectroscopic equipment, but is intended to give a survey on modern instrumentation used in laser spectroscopy.

There are several excellent and more detailed presentations of special instruments and spectroscopic techniques, such as spectrometers, interferometry, and Fourier spectroscopy. Besides the references given in the various sections, several series on optics [115], optical engineering [114], advanced optical techniques [284], and the monographs [117, 119, 283–288] may help to give more extensive information about special problems. Useful practical hints can be found in the handbooks [289, 290].

4.7 Problems

4.1 Calculate the spectral resolution of a grating spectrometer with an entrance slit width of 10 µm, focal lengths $f_1 = f_2 = 2\,\mathrm{m}$ of the mirrors M_1 and M_2, a grating with 1800 grooves/mm and an angle of incidence $\alpha = 45°$. What is the useful minimum slit width if the size of grating is $100 \times 100\,\mathrm{mm}^2$?

4.2 The spectrometer in Problem 4.1 shall be used in first order for a wavelength range around 500 nm. What is the optimum blaze angle, if the geometry of the spectrometer allows an angle of incidence α about 20°?

4.3 Calculate the number of grooves/mm for a Littrow grating for a 25° incidence at $\lambda = 488\,\mathrm{nm}$ (i.e., the first diffraction order is being reflected back into the incident beam at an angle $\alpha = 25°$ to the grating normal).

4.4 A prism can be used for expansion of a laser beam if the incident beam is nearly parallel to the prism surface. Calculate the angle of incidence α for which a HeNe laser beam ($\lambda = 632.8\,\mathrm{nm}$) transmitted through a rectangular flint glass prism with $\epsilon = 60°$ is expanded tenfold.

4.5 Assume that a signal-to-noise ratio of 50 has been achieved in measuring the fringe pattern of a Michelson interferometer with one continuously moving mirror. Estimate the minimum path length ΔL that the mirror has to travel in order to reach an accuracy of 10^{-4} nm in the measurement of a laser wavelength at $\lambda = 600\,\mathrm{nm}$.

4.6 The dielectric coatings of each plate of a Fabry–Perot interferometer have the following specifications: $R = 0.98$, $A = 0.3\,\%$. The flatness of the surfaces is $\lambda/100$ at $\lambda = 500\,\mathrm{nm}$. Estimate the finesse, the maximum transmission, and the spectral resolution of the FPI for a plate separation of 5 mm.

4.7 A fluorescence spectrum shall be measured with a spectral resolution of 10^{-2} nm. The experimentor decides to use a crossed arrangement of grating spectrometer (linear dispersion: 0.5 nm/mm) and FPI of Problem 4.6. Estimate the optimum combination of spectrometer slit width and FPI plate separation.

4.8 An interference filter shall be designed with peak transmission at $\lambda = 550\,\mathrm{nm}$ and a bandwidth of 5 nm. Estimate the reflectivity R of the dielectric coatings and the thickness of the etalon, if no further transmission maximum is allowed between 350 and 750 nm.

4.9 A confocal FPI shall be used as optical spectrum analyzer, with a free spectral range of 3 GHz. Calculate the mirror separation d and the finesse that is necessary to resolve spectral features in the laser output within 10 MHz. What is the minimum reflectivity R of the mirrors, if the surface finesse is 500?

4.10 Calculate the transmission peaks of a Lyot filter with two plates ($d_1 = 1$ mm, $d_2 = 4$ mm) with $n = 1.40$ in the fast axis and $n = 1.45$ in the slow axis (a) as a function of λ for $\alpha = 45°$ in (4.97); and (b) as a function of α for a fixed wavelength λ. What is the contrast of the transmitted intensity $I(\alpha)$ for arbitrary values of λ if the absorption losses are 2 %?

4.11 Derive (4.116) for the equivalent electrical circuit of Fig. 4.79b.

4.12 A thermal detector has a heat capacity $H = 10^{-8}$ J/K and a thermal conductivity to a heat sink of $G = 10^{-9}$ W/K. What is the temperature increase ΔT for 10^{-9} W incident cw radiation if the efficiency $\beta = 0.8$? If the radiation is switched on at a time $t = 0$, how long does it take before the detector reaches a temperature increase $\Delta T(t) = 0.9\Delta T_\infty$? What is the time constant of the detector and at which modulation frequency Ω of the incident radiation has the response decreased to 0.5 of its dc value?

4.13 A bolometer is operated at the temperature $T = 8$ K between superconducting and normal conducting states, where $R = 10^{-3}\,\Omega$. The heat capacity is $H = 10^{-8}$ J/K and the dc electrical current 1 mA. What is the change Δi of the heating current in order to keep the temperature constant when the bolometer is irradiated with 10^{-10} W?

4.14 The anode of a photomultiplier tube is connected by a resistor of $R = 1\,\text{k}\Omega$ to ground. The stray capacitance is 10 pf, the current amplification 10^6, and the anode rise time 1.5 ns. What is the peak amplitude and the halfwidth of the anode output pulse produced by a single photoelectron? What is the dc output current produced by 10^{-12} W cw radiation at $\lambda = 500$ nm, if the quantum efficiency of the cathode is $\eta = 0.2$ and the anode resistor $R = 10^6\,\Omega$? Estimate the necessary voltage amplification of a preamplifier (a) to produce 1 V pulses for single-photon counting; and (b) to read 1 V on a dc meter of the cw radiation?

4.15 A manufacturer of a two-stage optical image intensifier states that incident intensities of 10^{-17} W at $\lambda = 500$ nm can still be "seen" on the phosphor screen of the output state. Estimate the minimum intensity amplification, if the quantum efficiency of the cathodes and the conversion efficiency of the phosphor screens are both 0.2 and the collection efficiency of light emitted by the phosphor screens is 0.1. The human eye needs at least 20 photons/s to observe a signal.

4.16 Estimate the maximum output voltage of an open photovoltaic detector at room temperature under 10 μW irradiation when the photocurrent of the shortened output is 50 μA and the dark current is 50 nA.