



Practical OPTICS

Practical Optics

This page intentionally left blank

Practical Optics

Naftaly Menn



AmsterdamBostonHeidelbergLondonNew YorkOxfordParisSan DiegoSan FranciscoSingaporeSydneyTokyo

Elsevier Academic Press 200 Wheeler Road, 6th Floor, Burlington, MA 01803, USA 525 B Street, Suite 1900, San Diego, California 92101-4495, USA 84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper. 💿

Copyright © 2004, Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone: (+44) 1865 843830, fax: (+44) 1865 853333, e-mail: permissions@elsevier.com.uk. You may also complete your request on-line via the Elsevier homepage (http://elsevier.com), by selecting "Customer Support" and then "Obtaining Permissions."

Library of Congress Cataloging-in-Publication Data

Menn, Naftaly. Practical optics / Naftaly Menn. p. cm. Includes bibliographical references and index. ISBN 0-12-490951-5 (casebound : alk. paper) 1. Optics. I. Title. TA1520.M44 2004 621.36–dc22

2004005695

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN: 0-12-490951-5

For all information on all Academic Press publications visit our Web site at www.academicpress.com

Printed in the United States of America 04 05 06 07 08 09 9 8 7 6 5 4 3 2 1 To my children and grandchildren and in memory of my parents. This page intentionally left blank

Contents

Preface			xi			
1.	Geometrical Optics in the Paraxial Area					
	1.1.	Ray Opt	ics Conventions and Practical Rules.			
	Real and Virtual Objects and Images					
	1.2. Thin Lenses Layout. Microscope and Telescope Optical					
	Configurations					
	1.3. Diaphragms in Optical Systems. Calculation of Aperture Ang					
	and Field of View. Vignetting					
	1.4. Prisms in Optical Systems					
	1.5.	Solution	is to Problems	25		
2.	Theory of Imaging			41		
	2.1.	Optical A	Aberrations	41		
		2.1.1.	General Consideration. Ray Fan and Aberration Plot.			
			Concept of Wave Aberrations	41		
		2.1.2.	Chromatic Aberrations: Principles of Achromatic Lens			
			Design	44		
		2.1.3.	Spherical Aberration and Coma	48		
		2.1.4.	Aberrations of Tilted Beams (Field Aberrations)	51		
				vii		

		2.1.5. Sine Condition and Aplanatic Points	56	
		2.1.6. Addition of Aberrations	58	
	2.2.	Diffraction Effects and Resolution		
		2.2.1. General Considerations	61	
		2.2.2. Diffraction Theory of Imaging in a Microscope	64	
	2.3.	Image Evaluation		
	2.4.	Two Special Cases	71	
		2.4.1. Telecentric Imaging System	71	
		2.4.2. Telephoto Lens	72	
	2.5.	Solutions to Problems	73	
3.	Sources of Light and Illumination Systems			
	3.1.	Thermal Radiation Sources for Visible and IR	95	
	3.2.	Lens-based Illumination Systems	97	
	3.3.	Lasers	100	
		3.3.1. Main Characteristics of a Laser Beam	100	
		3.3.2. Beam Expansion and Spatial Filtering	104	
		3.3.3. Laser Diodes	108	
	3.4.	Light Emitting Diodes (LEDs)	112	
	3.5.	Solutions to Problems	114	
4.	Detectors of Light		129	
	4.1.	Classification of Radiation Detectors and Performance		
		Characteristics	129	
	4.2.	Noise Consideration	132	
	4.3.	Single Electro-optical Detectors (Photocells, Photomultipliers,		
		Semiconductor Detectors, Bolometers)	136	
	4.4.	Detector Arrays (One-dimensional Arrays and CCD and		
		CMOS Area Sensors)	143	
	4.5.	Solutions to Problems	148	
5.	Optical Systems for Spectral Measurements			
	5.1.	Spectral Properties of Materials and Spectral Instruments	159	
	5.2.	Prism-based Systems	172	
	5.3.	Diffraction Gratings and Grating-based Systems	178	
		5.3.1. Plane Diffraction Gratings and Related Configurations	178	

	Contents		
		5.3.2. Systems with Concave Diffraction Gratings	184
	5.4.	Interferometry-based Spectral Instruments	186
		5.4.1. Interference Filters and Fabry–Perot Interferometer	186
		5.4.2. Fourier Spectrometer	190
	5.5.	Spectrophotometry	193
	5.6.	Solutions to Problems	195
6.	Non	-contact Measurement of Temperature	209
	6.1.	Thermal Radiation Laws and Surface Properties	209
	6.2.	Optical Methods of Temperature Measurement	214
	6.3.	Measurement of Temperature Gradients	218
	6.4.	Solutions to Problems	221
7.	Opti	cal Scanners and Acousto-optics	229
	7.1.	Electro-mechanical Scanners	229
	7.2.	Acousto-optics and Acousto-optical Scanners	233
		7.2.1. Acousto-optical Effect and Acousto-optical Cell (AOM)	233
		7.2.2. Two Operation Modes: AOM as Modulator of Light	
		and AOM as Deflector of Optical Beams	237
		7.2.3. AOM Architecture for Spectral Analysis	239
	7.3.	Solutions to Problems	241
8.	Opti	cal Systems for Distance and Size Measurements	251
	8.1.	Laser Rangefinders	251
	8.2.	Size Measurement with a Laser Scanner	253
	8.3.	Interferometric Configuration	254
	8.4.	Stratified Light Beam and Imaging Measuring Technique	257
	8.5.	Solutions to Problems	259
9.	Opti	cal Systems for Flow Parameter Measurement	269
	9.1.	Principles of Laser Doppler Velocimetry (LDV)	269
	9.2.	Measurement of Velocity in 2-D and 3-D Flow Geometry	276
	9.3.	Two-phase Flow and Principles of Particle Sizing	281
	9.4.	Solutions to Problems	284

Contents

10.	Color and its Measurement			293	
	10.1.	Colo	or Sensation, Color Coordinates, and Photometric Units	293	
	10.2.	Colo	olor Detection and Measurement		
	10.3.	Solutions to Problems	304		
References					
Арре	Appendices				
A	Appendi	ix 1.	Physical Constants	311	
A	Appendix 2. Appendix 3.		Selected Data for Schott Optical Glasses	312	
A			Black Body Radiation	313	
A	Appendi	ix 4.	Emissivity of Selected Materials	315	
Inde	Index				

х

This book is intended primarily for students specializing in electro-optics. However, it is hoped that it will also serve engineers and R&D professionals, who, while not engaged directly in optics, are nevertheless involved with any of the numerous applications of optical methods of analysis.

The growing interest in electro-optics during last three decades has been accompanied by a rich literature, enlightening students and professionals in many topics of this wide-ranging field. So, why one more book on electro-optics? I have been involved in optics for more than 30 years, in both research and development and teaching of optical engineering. My deep involvement in the field has brought to my attention the gap between the theoretical study of optics and the ability of students and young engineers to solve practical problems arising in different applications. In this book I have tried to close this gap. For this reason the main focus I put is on solving the problems encountered in a variety of engineering and scientific applications.

All problems are original. Most of them serve not only for teaching purposes, but also provide useful information on specific applications: optical configuration for automatic inspection in industry, surveying systems, robot navigation, X-ray imaging, computerized radiography, microscopy vision and measurements, laser Doppler technique and flow study, non-contact measurement of temperature, acousto-optical scanners, spectral analysis, and many others. The solutions of problems are very detailed and include not only the theoretical approach and assumptions, but also the calculation procedure and units of measurements. Each chapter starts with the theoretical background related to the topic. The background comprises all relevant information and formulas required for the solution of the problems. In this regard the book is self-contained and very seldom it is necessary to consult additional references (these cases are clearly indicated in the text). Obviously, such an approach does not allow for detailed explanations of theoretical material or demonstration of the derivation procedure, but it makes the studying process easier, and, in my opinion, more effective.

The structure of the book reflects my understanding of the basic required knowledge in the field of electro-optics. The material in the book covers the theory of imaging, including geometrical optics, aberration theory and aspects of physical optics, a description of radiation sources and radiation detectors, spectroscopy systems, color measurements, and optical systems from different application areas. What is not included in the book is waveguide optics and communication systems – both topics are extensively covered in the existing literature.

I would be grateful for any comments, either related to the book's structure or to the solution of a specific problem.

The material included in this book served as a basis for the two-semester course on optical engineering which I have been teaching for a number of years. In teaching this course I was helped by many assistants to whom I am very much obliged. I am also grateful to my colleagues participating in numerous projects in industrial, military, and medical fields confronting me with problems many of which are included in this book. My special gratitude is to my wife, Irene, for her assistance in this work and for her endless patience, without which this book could not have been written.

> *Naftaly Menn* December 2003

Chapter 1

Geometrical Optics in the Paraxial Area

1.1. Ray Optics Conventions and Practical Rules. Real and Virtual Objects and Images

Electro-optical systems are intended for the transfer and transformation of radiant energy. They consist of active and passive elements and sub-systems. In active elements, like radiation sources and radiation sensors, conversion of energy takes place (radiant energy is converted into electrical energy and vice versa, chemical energy is converted in radiation and vice versa, etc.). Passive elements (like mirrors, lenses, prisms, etc.) do not convert energy, but affect the spatial distribution of radiation. Passive elements of electro-optical systems are frequently termed optical systems.

Following this terminology, an optical system itself does not perform any transformation of radiation into other kinds of energy, but is aimed primarily at changing the spatial distribution of radiant energy propagated in space. Sometimes only concentration of radiation somewhere in space is required (like in the systems for medical treatment of tissues or systems for material processing of fabricated parts). In other cases the ability of optics to create light distribution similar in some way to the light intensity profile of an "object" is exploited. Such a procedure is called imaging and the corresponding optical system is addressed as an imaging optical system.

Of all the passive optical elements (prisms, mirrors, filters, lenses, etc.) lenses are usually our main concern. It is lenses that allow one to concentrate optical energy or to get a specific distribution of light energy at different points in space (in other words, to create an "image"). In most cases experienced in practice,



FIGURE 1.1 Optical beams: (a) parallel, (b,c) homocentric and (d) non-homocentric.

imaging systems are based on lenses (exceptions are the imaging systems with curved mirrors).

The functioning of any optical element, as well as the whole system, can be described either in terms of ray optics or in terms of wave optics. The first case is usually called the geometrical optics approach while the second is called physical optics. In reality there are many situations when we need both (for example, in image quality evaluation, see Chapter 2). But, since each approach has advantages and disadvantages in practical use, it is important to know where and how to exploit each one in order to minimize the complexity of consideration and to avoid wasting time and effort.

This chapter is related to geometrical optics, or, more specifically, to ray optics. Actually an optical ray is a mathematical simplification: it is a line with no thickness. In reality optical beams which consist of an endless quantity of optical rays are created and transferred by electro-optical systems. Naturally, there exist three kinds of optical beams: parallel, divergent, and convergent (see Fig. 1.1). If a beam, either divergent or convergent, has a single point of intersection of all optical rays it is called a homocentric beam (Fig. 1.1b,c). An example of a non-homocentric beam is shown in Fig. 1.1d. Such a convergent beam could be the result of different phenomena occurring in optical systems (see Chapter 2 for more details).

Ray optics is primarily based on two simple physical laws: the law of reflection and the law of refraction. Both are applicable when a light beam is incident on a surface separating two optical media, with two different indexes of refraction, n_1 and n_2 (see Fig. 1.2). The first law is just a statement that the incident angle, *i*, is



FIGURE 1.2 Reflection and refraction of radiation.

equal to the reflection angle, i'. The second law defines the relation between the incident angle and the angle of refraction, r:

$$\sin(i)/\sin(r) = n_2/n_1.$$
 (1.1)

It is important to mention that all angles are measured from the vertical line perpendicular to the surface at the point of incidence (so that the normal incidence of light means that i = i' = r = 0).

In the geometrical optics approach the following assumptions are conventionally accepted:

- (a) radiation is propagated along a straight line trajectory (this means that diffraction effects are not taken into account);
- (b) if two beams intersect each other in space there is no interaction between them and each one is propagated as if the second one does not appear (this means that interference effects are not taken into account);
- (c) ray tracing is invertable; in other words, if the ray trajectory is found while the ray is propagated through the system from input to output (say, from the left to the right) and then a new ray comes to the same system along the outgoing line of the first ray, but propagates in the reverse direction (from the right to the left), the trajectory of the second ray inside and outside of the system is identical to that of the first ray and it goes out of the system along the incident line of the first ray.

Normally an optical system is assumed to be axisymmetrical, with the optical axis going along OX in the horizontal direction. Objects and images are usually located in the planes perpendicular to the optical axes, meaning that they are along the OY (vertical) axis. Ray tracing is a procedure of calculating the trajectory of optical rays propagating through the system. Radiation propagates from the left to the right and, consequently, the object space (part of space where the light sources or the objects are located) is to the left of the system. The image space (part of space where the light detectors or images are located) is to the right of the system.

All relevant values describing optical systems can be positive or negative and obey the following sign conventions and rules:

- *ray angles* are calculated relative to the optical axis; the angle of a ray is positive if the ray should be rotated counterclockwise in order to coincide with OX, otherwise the angle is negative;
- vertical segments are positive above OX and negative below OX;
- *horizontal segments* should start from the optical system and end at the relevant point according to the segment definition. If going from the starting point



FIGURE 1.3 Sign conventions.

to the end we move left (against propagated radiation), the segment is negative; if we should move right (in the direction of propagated radiation), the corresponding segment is positive.

Examples are demonstrated in Fig. 1.3. The angle *u* is negative (clockwise rotation of the ray to OX) whereas *u'* is positive. The object Y is positive and its image Y' is negative. The segment *S* defines the object distance. It starts from the point O (from the system) and ends at the object (at Y). Since we move from O to Y against the light, this segment is negative (S < 0). Accordingly, the segment *S'* (distance to the image) starts from the system (point O') and ends at the image Y'. Since in this case we move in the direction of propagated light (from left to right) this segment is positive (S' > 0).

The procedure of imaging is based on the basic assumption that any object is considered as a collection of separate points, each one being the center of a homocentric divergent beam coming to the optical system. The optical system transfers all these beams, converting each one to a convergent beam concentrated in a small spot (ideally a point) which is considered as an image of the corresponding point of the object. The collection of such "point images" creates an image of the whole object (see Fig. 1.4).



FIGURE 1.4 Concept of image formation.

An ideal imaging is a procedure when all homocentric optical beams remain homocentric after traveling through the optical system, up to the image plane (this case is demonstrated in Fig. 1.4). Unfortunately, in real imaging the outgoing beams become non-homocentric which, of course, "spoils" the images and makes it impossible to reproduce the finest details of the object (this is like a situation when we try to draw a picture using a pencil which is not sharp enough and makes only thick lines – obviously we fail to draw the small and fine details on the picture). The reasons for such degradation in image quality lie partially in geometrical optics (then they are termed optical aberrations) and partially are due to the principal limitations of wave optics (diffraction limit). We consider this situation in detail in Chapter 2. Here we restrict ourselves to the simple case of ideal imaging.

In performing ray tracing one should be aware that doing it rigorously means going step by step from one optical surface to another and calculating at each step the incident and refraction angles using Eq. (1.1). Since many rays should be calculated, it is a time-consuming procedure which today is obviously done with the aid of computers and special programs for optical design. However, analytical consideration remains very difficult (if possible at all). The complexity of the procedure is caused mainly by the nonlinearity of the trigonometrical functions included in Eq. (1.1). The situation can be simplified drastically if we restrict ourselves to considering small angles of incidence and refraction. Then $\sin(i) \approx i$; $\sin(r) \approx r$; r = i/n and all relations become linear. Geometrically this approximation is valid only if the rays are propagated close to the optical axis of the system, and this is the reason why such an approximation is called paraxial. A paraxial consideration enables one to treat optical systems analytically. Because of this, it is very fruitful and usually is exploited as the first approximation at the early stage of design of an optical system.

Even in the paraxial approach we can further simplify the problem by neglecting the thickness of optical lenses. Each lens consisting of two refractive surfaces (spherical in most cases, but sometimes they could be aspherical) separated by glass (or other material) of thickness t is considered as a single "plane element" having no thickness, but still characterized by its ability to concentrate an incident parallel beam in a single point (called the focal point or just focus). In such a case the only parameter of the lens is its focal length, f', measured as the distance between the lens plane and the focus, F'. Each lens has two focuses: the back (F') and the front (F), the first being the point where the rays belonging to a parallel beam incident on the lens from the left are concentrated and the second being the center of the concentrated rays when a parallel beam comes to the lens from the right. Obviously, if the mediums at both sides of the lens are identical (for example, air on both sides or the lens being in water) then f' = -f. In the case when the mediums are different (having refractive index n and n' correspondingly) the relation should be

$$nf' = -n'f. \tag{1.2}$$

The optical power of a lens, defined as

$$\Phi = 1/f',\tag{1.3}$$

is used sometimes in system analysis, as we shall see later.

Imaging with a simple thin lens obeys the two following equations:

$$\frac{1}{S'} - \frac{1}{S} = \frac{1}{f'} \tag{1.4}$$

$$V = S'/S = y'/y$$
 (1.5)

where V is defined as the optical magnification. These two formulas enable one to calculate the positions and sizes of images created by any thin lens, either positive or negative, if all values are defined according to the sign conventions and rules described earlier in this section. A number of thin lenses which form a single system can also be treated using expressions (1.4) and (1.5) step-by-step for each component separately, the image of element *i* being considered as a virtual object for element (i + 1). An example of such a consideration with details for a two-lens system is presented in Problem 1.7.

The next step in approaching the real configuration of an optical system is to take into account the thickness of its optical elements. Still remaining in the paraxial range one can describe the behavior of a single spherical surface (see Fig. 1.5) by the Abbe invariant (r is the radius of the surface):

$$n\left(\frac{1}{r} - \frac{1}{S}\right) = n'\left(\frac{1}{r} - \frac{1}{S'}\right).$$
(1.6)



FIGURE 1.5 Refraction of rays at a single spherical surface.



FIGURE 1.6 Ray tracing between two spherical surfaces.

Then, the ray tracing for an arbitrary number of surfaces can be performed with the aid of the following two simple relations (see also Fig. 1.6):

$$u_{k+1} = \frac{n_k}{n_{k+1}} u_k + \frac{h_k}{r_k n_{k+1}} (n_{k+1} - n_k)$$
(1.7)

$$h_{k+1} = h_k - u_{k+1}d_k$$
 (k = 1, 2, ..., N). (1.8)

Given the radii of the spherical surfaces, the refraction indexes on both sides, and the distances between them, all angles, u_k , and heights, h_k , can be easily found, starting from initial values u_1 , h_1 .

To apply Eqs. (1.7) and (1.8) to a single lens defined by two spherical surfaces of radii r_1 and r_2 separated by the segment d, we first have to remind ourselves of the definition of the principal planes, H, H' and the cardinal points. As is seen from Fig. 1.7, the real ray trajectory ABCD can be replaced by ABMM'CD in such a way that they are identical outside the lens, but inside the lens the rays intersect two virtual planes H and H' at the same height (OM = O'M'). Actually these principal planes, H, H', can represent the lens as far as ray tracing is considered. Furthermore, the focal distances, f, f', are measured from the cardinal points O, O' to the front and back focuses, F and F'. The terms "back focal length" (BFL) and "front focal length" (FFL) are related to the segments $S_{F'}$, S_F from the back and front real surfaces to F' and F, respectively (see Fig. 1.7). Calculation of BFL and FFL enables one to determine the location of both principal planes with regard to the lens surfaces. Leaving the details of calculation to Problem 1.5 we just indicate here the final results:

$$S_{F'} = f' \left[1 - \frac{d}{r_1 n} (n-1) \right]$$
(1.9)

$$S_F = -f' \left[1 + \frac{d}{r_2 n} (n-1) \right]$$
(1.10)



FIGURE 1.7 Principal planes of a thick lens.

and for the focal distance

$$\frac{1}{f'} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) + \frac{d(n-1)^2}{r_1 r_2 n}.$$
(1.11)

In many cases the second term of the last formula can be neglected since it is much smaller than the first one.

Problems

1.1. Find the image of the object OA in Fig. 1.8 using the graphical method.



FIGURE 1.8 Problem 1.1 – Imaging by the graphical method.

1.2. Find the image of the point source A and direction of the ray AB after the positive lens L_1 (Fig. 1.9a) and the negative lens L_2 (Fig. 1.9b).

1.3. Ray tracing in a system of thin lenses. Find the final image of a point source A after an optical system consisting of thin lenses L₁, L₂, and L₃ ($f'_1 = f'_2 = 15 \text{ mm}$; $f'_3 = 20 \text{ mm}$) if A is located on the optical axis 30 mm left of the lens L₁



FIGURE 1.9 Problem 1.2 – Imaging by the graphical method: (a) with a positive lens; (b) with a negative lens.

and the distances between the lenses are $d_{12} = 40$ mm, $d_{23} = 60$ mm. [Note: Do this by ray tracing based on Eq. (1.4).]

1.4. Method of measurement of focal length of a positive lens. An image of an object AB created by a lens is displayed on a screen P distant from AB at L = 135 mm (Fig. 1.10). Then the lens is moved from the initial position, 1, where the sharp image is observed at magnification V_1 , to the position 2 where again the sharp image is observed on the same screen, but at magnification $V_2 = 1/V_1$. The distance between positions 1 and 2 is a = 45 mm. Find the focal length of the lens and estimate the uncertainty of the measured value if the lens thickness, *t*, is about 5–6 mm.



FIGURE 1.10 Problem 1.4 – Method of focal length measurement.

1.5. Location of the principal planes of a thick lens. Find the positions of two principal planes H and H', BFL, and FFL of a lens made of glass BK-7 (n = 1.5163) having two spherical surfaces of radii $R_1 = 50$ mm and $R_2 = -75$ mm and thickness t = 6 mm.



FIGURE 1.11 Problem 1.6 – Consideration of a parallel plate.

1.6. Violation of homocentricity of a beam passed through a flat slab. A flat slab of glass is illuminated by a homocentric beam which fills the solid angle $\omega = 1.5$ sr with the center at point A, 30 mm behind the slab (Fig. 1.11). The thickness of the slab t = 5 mm and refractive index n = 1.5. Find the location of the point A' after the slab as a function of incident angle, *i*, and estimate the deviation of the outgoing beam from homocentricity.

1.7. A two-lens system in the paraxial range. A lens L_1 of 100 mm focal length is followed by a lens L_2 of 75 mm focal length located 30 mm behind it. Considering both lenses as a unified system find the equivalent optical power and position of the focal plane.

1.8. A ball lens. Find the location of the principal planes of a ball lens (a full sphere) of radius r = 3 mm and its BFL.

1.2. Thin Lenses Layout. Microscope and Telescope Optical Configurations

We will consider here the following basic configurations: (i) magnifier; (ii) microscope; and (iii) telescope. All three can be ended either by a human eye or by an electro-optical sensor (like a CCD or other area sensor).

The Human Eye

Although the details of physiological optics are beyond the scope of this book, we have to consider some important features of the human eye (for further details, see Hopkins, 1962) as well as eye-related characteristics of optical devices. Usually the "standard eye" (normal eye of an adult person) is described in terms of a simplified model (so-called "reduced eye") as a single lens surrounded by the air

from the outside, and by the optically transparent medium (vitreous humor) of refractive index 1.336 from the inside. As a result, the front focal length of the eye, f, differs from the back focal length, f' (see Eq. (1.2)). The front focal length is usually estimated as 17.1 mm whereas f' is equal to 22.9 mm. The pupil of the eye varies from 2 mm (minimum size) to 8 mm (maximum size) according to the scene illumination level (adaptation). The lens creates images on the retina which consists of huge numbers of photosensitive cells. The average size of the retina cells dictates the angular resolution of the eye (ability of seeing two small details of the object separately). The limiting situation is that the images of two points are created at two adjacent cells of the retina. This renders the angular resolution of a normal eye to be 1 arcminute (3 \times 10⁻⁴ rad). The lens curvature is controlled by the eye muscles in such a way that the best (sharp) image is always created on the retina, whether the object is far or close to the eye (accommodation process). The distance of best vision is estimated as 250 mm, which means that the eye focused on objects at distances of 250 mm is not fatigued during long visual operation and can still differentiate small details.

Three kinds of abnormality of eye optics are usually considered: myopia, hyperopia, and astigmatism. The first one (also called near-sightedness) occurs when a distant object image is not created on the retina but in front of it. A corrective negative lens is required in such a situation. In the second case (called far-sightedness) the opposite situation takes place: the images are formed behind the retina and, obviously, the corrective lens should be positive. Astigmatism means that the lens curvature is not the same in different directions which results in differences in focal lengths, say in the horizontal and vertical planes. Correction is done by spectacles with appropriately oriented cylindrical lenses.

The other properties of the eye related to visual perception are considered in Chapter 10.

Magnifications in Optical Systems

Generally, four adjacent magnifications can be defined for any optical system: (i) linear magnification, V, for objects and images perpendicular to the optical axis; (ii) angular magnification, W; (iii) longitudinal magnification, Q (magnification in the direction of the optical axis); and (iv) visible magnification, Γ (used only for systems working with the human eye).

Linear magnification, defined earlier for a single lens by Eq. (1.5), is still applicable for any complete optical system. Angular magnification can be defined for any separate ray or for a whole beam incident on a system. For example, for the tilted ray shown in Fig. 1.12 *W* is calculated as follows:

$$W = \tan(u')/\tan(u). \tag{1.12}$$



FIGURE 1.12 Imaging of vertical and horizontal segments.



FIGURE 1.13 Explanation of visible magnification.

As can be shown, the product *VW* is a system invariant: it does not depend on the position of the object and image, but is determined by the refractive indexes on both sides of the optical system (*n* and *n'*). If n = n' then VW = 1.

Considering the segment l along the optical axis and two pairs of conjugate points, A and A', C and C' (Fig. 1.12), we can find the longitudinal magnification, Q:

$$Q = l'/l. \tag{1.13}$$

It can be shown that for small segments l, l' one can use the formula $Q = V^2$.

Finally, visible magnification is related to the size of images on the retina of an eye. It is defined as the ratio of the image created by the optical system to the image of the same object observed by the naked eye directly. Since the image size is proportional to the observation angle (see Fig. 1.13), Γ is determined as follows:

$$\Gamma = \tan(\gamma')/\tan(\gamma). \tag{1.14}$$

A Simple Magnifier

This is usually operated with the eye. While observing through a magnifying glass an object is positioned between the front focus of the lens and the lens



FIGURE 1.14 A simple magnifier.

itself (Fig. 1.14). The image is virtual and its position corresponds to the distance of best vision of the eye (250 mm). The closer the object to F, the higher the magnification. Therefore, approximately, we can define that $s \approx f$ (of course s < 0; f < 0), and for visible magnification of the magnifier we have

$$\Gamma = \frac{250}{f'}.\tag{1.15}$$

Since the distance of best vision is much greater than the focal distance of the eye, the rays coming to the eye pupil are almost parallel. In most cases they can be treated just as a parallel beam (or beams).

The Microscope

Figure 1.15a demonstrates the basic layout of a microscope working with the eye and Fig. 1.15b shows a microscope working with an electro-optical detector (like a CCD or other video sensor). In both cases the first lens L₁ (called the objective) is a short-focus well-corrected lens creating the first real magnified image of the object (AB) in the plane P. The second lens is the eyepiece (Fig. 1.15a) or the relay lens (Fig. 1.15b). The eyepiece L₂ functions like a simple magnifier and its visible magnification obeys Eq. (1.15). Magnification of the objective, V_1 , can be found from Eq. (1.5). Usually the object distance S_1 is very close to f_1 and the distance $S'_1 = T$ from the lens L₁ to the plane P is chosen as one of several standardized values accepted by all manufacturers of microscopes (160 mm or 180 mm or 210 mm). Therefore, for the total magnification of the microscope working with eye we get:

$$V_{\rm M} = V_1 V_2 = \frac{T}{f_1'} \times \frac{250}{f_2'}.$$
 (1.16)







FIGURE 1.15 Layout of a microscope working with (a) the eye and (b) an area detector.

If instead of an eyepiece a relay lens is exploited its actual linear magnification, V_2 , should be taken into account, and then

$$V_{\rm M} = \frac{T}{f_1'} V_2. \tag{1.17}$$

If the microscope is intended for measurements and not only for observation then a glass slab with a special scale (a ruler, a crosshair, etc.) called a reticle is introduced in the plane P. In such a case the eye observes the image overlapped with the scale.

In Fig. 1.15 the rays originating from two points of the object are drawn – from the center of the object (point O) and from the side (point A). As can be seen from Fig. 1.15a, each point gives a parallel beam after the eyepiece: one is parallel to the optical axis and the other is tilted to OX. Intersection of the beams occurs in the plane M (exit pupil of the microscope) where the operator's eye should be positioned.

For the convenience of the operator the optical layout in most cases is split after the plane P in two branches, each one having a separate eyepiece. Such an output assembly is called binocular and observation is done by two eyes. It should be understood, however, that binocular itself does not render stereoscopic vision, since both eyes are observing the same image created by a single objective L_1 . To achieve a real stereoscopic effect two objectives are required in order to observe the



FIGURE 1.16 Microscope with a trinocular assembly.



FIGURE 1.17 Layout of a microscope with ICS optics.

object from two different directions. Each image is transferred through a separate branch (a pair of lenses L_1 and L_2).

The architecture shown in Fig. 1.16 is actually the combination of the two layouts presented in Fig. 1.15 and its output assembly is called trinocular – it creates images on the area sensor as well as in the image plane of both eyepieces. The beam splitter, BS, turns the optical axis in the direction of the relay lens L_3 .

In the last few years microscopes have been designed as infinity color-corrected systems (ICS) which means that the object is located in the front focal plane of the objective, its image is projected to infinity, and an additional lens L_4 (the tube lens) is required in order to create an intermediate image in the plane P. Such a layout is demonstrated in Fig. 1.17. One of the important advantages of ICS optics is that the light beams are parallel between L_1 and L_4 enabling one to introduce here optical filters with no degradation of the optical quality and without relocation of the image plane P.

The Telescope

Telescopic systems are intended for observation of remote objects. If the distance between the object and the first lens of the system is much greater than the focal length of L_1 all light beams at the entrance of the system can be considered as

parallel, whether they are coming from the central point of the object or from the side.

Again, as in the above consideration of a microscope, the central point beam is parallel to the optical axis whereas the side point generates an oblique parallel beam. All incident beams are concentrated by the objective in its back focal plane (passing through the back focus F'_1). The second lens L_2 is positioned in such a way that its front focus F_2 coincides with F'_1 . Obviously all beams after lens L_2 become parallel again, but the exit angles of the oblique rays are different from those of the corresponding beams at the entrance (see Fig. 1.18) and this causes the angular magnification of the telescope to be (see Eq. (1.12))

$$W = \tan(\beta')/\tan(\beta) = f_1'/f_2'.$$
 (1.18)

As follows from Eq. (1.18), the longer the focal length of the objective, the greater the magnification. However, along with this the necessary size of the lens L₂ also increases, which might cause a limitation of the visible field of view (the part of the object space visible through the system). To solve this problem an additional lens L₃ (the field lens) is introduced in the system (see Fig. 1.19). This lens allows one to vary the vertical location of the oblique beam incident on L₂.

The configurations shown in Figs. 1.18 and 1.19 are built of positive lenses. In the Galilean architecture the eyepiece L_2 is negative (Fig. 1.20). As a result,



FIGURE 1.18 Basic layout of a telescope.



FIGURE 1.19 Telescope with a field lens.



FIGURE 1.20 Galilean telescope.

the total length of the system is shortened. However, the intermediate image is virtual (both focal points F'_1 and F_2 are behind the eyepiece) and there is no way to introduce a measurement scale, if necessary. However, as it turns out, this shortcoming becomes very useful if the Galilean configuration is exploited with high-power lasers (for beam expanding).

Problems

1.9. If the angular resolution of the eye is 3×10^{-4} rad, what is the average size of the retina cells?

1.10. A microscope is intended for imaging an object located in the plane P simultaneously in two branches: one for observation by eye and the other for imaging onto a plane area sensor (CCD). The objective of the microscope serving the two branches is of 20 mm focal length and provides linear magnification V = -10 to the image plane of the eyepiece where a reticle M of 19 mm diameter is positioned (Fig. 1.21). The CCD sensor is 4.8 mm (vertical) × 5.6 mm (horizontal)



FIGURE 1.21 Problem 1.10 – Two-branch microscope.

in size. In front of the CCD at a distance of 20 mm an additional relay lens L_3 is introduced in order to reach the best compatibility of the field of view in both branches. Assuming the eyepiece L_2 to be of 25 mm focal length and neglecting the thickness of the lenses, find:

- (a) the working distance (location of the object plane P with regard to the objective);
- (b) the total magnification in the branch to the eye;
- (c) the optical power of lens L₃. [Note: Find two solutions and choose the one which provides the shortest distance between P and the CCD.]

1.11. Dual magnification system with negative relay lens. Such a system is widely used in the microelectronics industry where automatic processing of wafers is a main concern. The object (usually a wafer) is located in the plane P and imaged onto a CCD either at low magnification $V_1 = -3$ (through the right branch of the arrangement, exploited for initial alignment) or at high magnification $V_2 = 2 \times V_1$ (fine alignment through the left branch where lens L_2 and retroreflector R are introduced). While switching the system between two alignment procedures no optical element should be moved, except the aperture D (Fig. 1.22). The retroreflector R allows one to vary the high magnification of the system with minimum effort – just replacement of R and L_2 , with no other changes in the arrangement. Thus, lens L_2 serves as a negative relay lens of the system. Neglecting the thickness of the lenses and taking all necessary distances from Fig. 1.22, find:

- (a) the focal length of L₂ and its position with regard to the CCD and the other elements of the arrangement;
- (b) the relocation of the retroreflector and the relay lens L_2 from their initial positions required to increase the magnification in the left branch by 10%.



FIGURE 1.22 A dual magnification system.

1.3. Diaphragms in Optical Systems. Calculation of Aperture Angle and Field of View. Vignetting

The size of each optical element of a system should be considered properly, since it influences: (i) the quantity of radiant energy passing through the system; (ii) the quality of images; and (iii) the cost of the system. Among all the geometrical parameters the working diameter is of primary importance (remember that we assume that the system is rotationally symmetric) – it acts as the transparent part of the element.

Sometimes an additional diaphragm (a physical element called a stop which has a final size aperture and negligible thickness) is introduced in the system. An aperture stop is a diaphragm which actually limits the size of light bundles passing through the system and consequently it is responsible for the amount of energy collected at each point of the image. The aperture stop is illustrated in Fig. 1.23. Assume that the system consists of a number of elements (of which the first and the last curved surfaces are shown in the figure) and also includes the stop cd. The boundaries of each optical surface are also considered as diaphragms. First we "transfer" all the diaphragms into the object space (e.g., we find the size and location of the image of each diaphragm through the rest of the optical elements to the left of it, as if the light beams are propagated from right to left). Such an image of the stop cd is c'd'; the image of the first diaphragm ab is ab itself, since there is no element left of it; the third diaphragm shown in the figure is the image of some other optical surface, etc. Then we connect the ray from the central point O of the object to the side of each image and find the angle of each ray with the optical axis. The smallest angle (in our example it is the angle of the ray Oc') is called the aperture angle, α_{ap} , and the corresponding physical diaphragm is called the aperture stop (cd in the case of Fig. 1.23). Its image in the object space is called



FIGURE 1.23 Aperture stop and entrance and exit pupils.

the entrance pupil and its image to the image space is called the exit pupil (c'd' and c''d'', respectively). Obviously, the aperture angle defines the maximum cone of light rays emerging from point O and passing through the system with no obstacle up to point O' in the image plane. The corresponding angle α'_{ap} is the aperture angle in the image space. Drawing the rays that connect any other point of the object with the entrance pupil we find the corresponding cone of light participating in imaging of that point. The ray connecting the oblique point A with the center C of the entrance pupil is called the chief ray (shown by the dotted line in Fig. 1.23). Its position in the image space is C'A'.

Now we consider the entrance pupil, ab, together with any other diaphragm (or its image, gh) in the object space (Fig. 1.24). It is understood that the conical bundle originating from point O of the object is not affected by gh at all. The same is true for any other point of the object plane between O and A where the last one is found with the ray passing through the sides a and g of both diaphragms. For remote points above A (point B, for example), the light cone filling the entrance pupil is cut partially by the diaphragm gh (the dotted line originating in B cannot be transferred). This means that the active cone of light passing through the system is reduced gradually until we achieve finally the point C from which no ray can pass the system. The rays emerging from any point above C cannot achieve the image plane at all. Therefore, image formation can be performed only for a part of the object plane (the circle of radius OC). This part of the object plane is called the field of view and the diaphragm gh is called the field aperture. If gh is the image of a real physical diaphragm GH located somewhere in the system then GH is called the field stop.

Reduction of the light cones while moving out from the optical axis causes a decrease of the image brightness in the corresponding parts of the image plane. Even if the object plane is equally illuminated we get a reduction of the brightness in the image plane, as is illustrated by the graph of intensity, I(r), in Fig. 1.24.



FIGURE 1.24 Field of view and vignetting.



FIGURE 1.25 Finding the field aperture.

This phenomenon is known as vignetting, and it should be carefully investigated if a new optical system is designed.

To find the field aperture it is necessary to image all physical diaphragms of the system into the object space, to calculate the sizes and location of each image, and then to draw a ray connecting the center of the entrance pupil with the side point of each image and to calculate the corresponding angle with the optical axis. The minimum angle defines the field aperture (and consequently the field stop). The procedure described is illustrated by Fig. 1.25 where g_2h_2 serves as the field aperture. It is useful to take into account the fact that to avoid vignetting it is necessary to position the field stop at the plane of the intermediate image of the system.

Problems

1.12. The system of two thin lenses L_1 (focal length 100 mm, diameter 20 mm) and L_2 (focal length 50 mm, diameter 20 mm) shown in Fig. 1.26 forms an image of the object plane P on a screen M at magnification V = 3. The distance between P and L_1 is 200 mm.



FIGURE 1.26 Problem 1.12 – Imaging with two lenses.

- (a) How can the field stop ab of the system be positioned in order to get imaging with no vignetting?
- (b) What should be the size of the field stop if the field of view is 10 mm?
- (c) Find the location of all elements of the system and calculate the aperture angle and position of the entrance pupil.

1.13. In the system of Problem 1.10, find the minimum size of lens L_3 which enables one to get images on the CCD with no vignetting.

1.4. Prisms in Optical Systems

Prisms serve three main purposes in optical systems: (i) to fold the optical axis; (ii) to invert the image; and (iii) to disperse light of different wavelengths. The latter is discussed in detail in Chapter 5. Here we will consider the first two purposes. It is quite understandable that both are achieved due to reflection of rays on one or several faces of the prism. So, it is worth keeping in mind how the system of plane reflectors (plane mirrors) can be treated (e.g., see Problem 1.14).

A great variety of prisms are commonly used in numerous optical architectures. Only a few simple cases are described below.

Right-angle prism. This is intended for changing the direction of the optical axis through 90°. The cross-section of this prism is shown in Fig. 1.27a. The rays coming from the object (the arrow 1–2) strike the input face AB at 90° and after reflection from the hypotenuse side emerge along the normal to the face BC. It can be seen that beyond the prism the object is inverted. The shortcoming of this prism is revealed if the incoming light is not normal to the prism face. In this case the angle between incoming and outgoing rays differs from 90°. Another issue is concerned with the total reflection of rays on face AC: it might happen that for



FIGURE 1.27 Layout of different prisms: (a) right-angle prism; (b) penta-prism; (c) Dove prism.



some tilted rays total reflection does not occur. In such a case a reflecting coating on AC is required.

Penta-prism. This prism has effectively four faces with an angle of 90° between AB and BC and 45° between the two other sides (Fig. 1.27b). The shortcoming of the right-angle prism does not occur here, i.e., the outgoing beam is always at 90° to the input beam, independent of the angle of incidence. Also, the object is not inverted. This results from a double reflection in the prism and is evidence of the common rule for any prism or system with reflectors; namely, the image is not inverted if the number of reflections is even.

Dove prism. The angles A and D are of 45° and the input and output beams are usually parallel to the basis face AD (Fig. 1.27c). While traveling through the prism the beams are inverted. Another feature of this prism is its ability to rotate an image: when the prism is inserted in an imaging system rotated around the input beam with angular speed ω the image in the system will be rotated at a speed 2ω .

If it is necessary to invert beams around two axes a combination of prisms, like the Amici prism shown in Fig. 1.28, can be exploited. This prism is actually a right-angle prism with an additional "roof" (for this reason it is also called the roof-prism). As a result the beams are inverted in both directions: upside-down and left–right.

In general, any prism inserted in an imaging system makes the optical path longer. This effect should be taken into account if a system designed for an unbent configuration has to be bent to a more compact size using prisms and mirrors. With regard to its influence on image quality and optical aberrations the prism acts as a block of glass with parallel faces. As was demonstrated earlier (see Problem 1.6 where the propagation of a divergent–convergent beam through a glass slab of thickness *t* was considered) the block of glass causes a lengthening of the optical path by (n - 1)t/n compared to the ray tracing in air. Therefore instead of tracing the rays through the slab and calculating the refraction at the entrance and exit


FIGURE 1.29 Unfolded diagram for (a) the right-angle prism, (b) the penta-prism, and (c) the Dove prism.

c)

surfaces one can replace a real plate by a virtual "air slab" of reduced thickness, t/n, and perform ray tracing for air only. To apply this approach to prisms we have to find the slab equivalent to the prism with regard to the ray path inside the glass. This can be done by the following procedure based on unfolded diagrams (see Fig. 1.29). We start moving along the incident ray until the first reflection occurs. Then we build the mirror image of the prism and the rays and proceed moving further along the initial direction until the second reflected surface is met. Then again we build the mirror image of the configuration, including the ray path, and proceed further until the initial ray leaves the last (exit) face of the prism. Details of the procedure can be seen in Problem 1.15.

Creating unfolded diagrams is aimed at calculating the thickness, t_e , of the equivalent glass block. For the cases depicted in Fig. 1.29:

- (a) right-angle prism with an entrance face of size *a*: $t_e = a$;
- (b) penta-prism with the same size *a* of the entrance face: $t_e = a(2 + \sqrt{2})$;
- (c) Dove prism of height *a* and 45° angles between faces: $t_e = 3.035a$.

Once t_e is known, the apparent thickness in air is calculated from t_e/n .



FIGURE 1.30 Problem 1.14 – Imaging in a mirror corner.

Problems

1.14. *Imaging in systems of plane mirrors.* An object AB is positioned as shown in Fig. 1.30 in front of a mirror corner of 45° . Find the location of the image beyond the mirrors.

1.15. Find the reduced (apparent) thickness of a 45° rhomboidal prism of 2 cm face length. The prism is made of BK-7 glass (n = 1.5163).

1.16. A lens L of 30 mm focal length transfers the image of an object AB positioned 40 mm in front of L to a screen P. A penta-prism with 10 mm face size is inserted 20 mm beyond the lens. Find the location of the screen P relative to the prism if it is made of BK-7 glass (n = 1.5163).

1.17. *Dispersive prism at minimum deviation.* Find the minimum deviation angle of a prism with vertex angle $\beta = 60^{\circ}$. The prism is made of SF-5 glass with refractive index n = 1.6727.

1.5. Solutions to Problems

1.1. We are looking for a solution in the paraxial range and assume the lens is of negligible thickness. To find the image of point A we use two rays emerging from A: ray 1 parallel to the optical axis and ray 2 passing through the center of the lens (Fig. 1.31). Ray 1 after passing through the lens goes through the back



FIGURE 1.31 Problem 1.1 – Graphical method of finding the image.



FIGURE 1.32 Problem 1.2 – Graphical method of finding the image with (a) a positive lens and (b) a negative lens.

focus F'. Ray 2 does not change its direction and continues beyond the lens along the incident line. The intersection of the two rays after the lens creates the image A' of point A. Once the image A' is found, the image O' of point O is obtained as the intersection of the normal from point A' to the optical axis.

It should be noted that instead of ray 1 or 2 one can use ray 3 (dotted line) going through the front focus F in the object space (in front of the lens) and parallel to OX after the lens. Intersection with the two other rays occurs, of course, at the same point A'. Also note that in our approximation of the paraxial range the homocentric beam also remains homocentric in the image space.

1.2. In both cases, Figs. 1.32a and b, we draw the ray (dotted line) parallel to AB and passing through the center of the lens. The ray crosses the back focal plane at point C. Since the ray and AB belong to the same parallel oblique bundle and all rays of such a bundle are collected by the lens in a single point of the back focal plane, this must be point C. Therefore, the ray AB after passing through the lens goes from B through C to point A' at the intersection with the axis. This point is the image of A. In the case of Fig. 1.32b the focus F' and corresponding back focal plane are located to the left of the lens. Hence, not the ray itself but its continuation passes through point C. The intersection with OX is still the image of the point source A which becomes virtual in this case.

1.3. First we will derive the ray tracing formula valid for the paraxial approximation. By multiplying both sides of Eq. (1.4) by *h* (see Fig. 1.33) and denoting $h/S = \tan(u) \approx u$ and $h/S' = \tan(u') \approx u'$ we get

$$u' - u = h\Phi$$

which yields for a number of lenses (k = 1, 2, ..., N):

$$u_{k+1} = u_k + h_k \Phi_k \tag{A}$$



FIGURE 1.33 Problem 1.3 – Ray tracing through a single lens.



FIGURE 1.34 Problem 1.3 – Ray tracing through a system of three lenses.

with the additional geometrical relation

$$h_{k+1} = h_k - u_{k+1} d_{k,k+1}.$$
 (B)

Expressions (A) and (B) enable one to calculate the ray trajectory in a system of thin lenses. To start the calculation we need the values u_1 , h_1 . Usually these values can be arbitrarily chosen, as they do not affect the final results. Going back to the numerical data of the problem, we choose $u_1 = -0.1$ and then proceed as follows (see Fig. 1.34):

$$u_{1} = -0.1; \quad h_{1} = S_{1}u_{1} = (-30)(-0.1) = 3.0$$

$$u_{2} = -0.1 + \frac{3.0}{15} = 0.1; \quad h_{2} = 3.0 - 0.1 \times 40 = -1.0$$

$$u_{3} = 0.1 - \frac{1.0}{15} = 0.0333; \quad h_{3} = -1.0 - 0.333 \times 60 = -3.0$$

$$u_{4} = 0.0333 - \frac{3.0}{20} = -0.1167; \quad S'_{3} = \frac{h_{3}}{u_{4}} = \frac{-3.0}{-0.1167} = 25.71 \text{ mm}.$$

It can be easily checked that exactly the same result will be obtained if we choose another initial value of u_1 (say, $u_1 = -0.2$). Of course, this results from the linearity of the expressions (A) and (B).

1.4. Measurement of the optical power of a lens or its focal length is often required in the practice of optical testing. The method described here is particularly useful because it is based only on measurements of linear distance (L) and linear displacement (a) which can be easily and accurately realized.

We start with a derivation of working formulas. Combining Eqs. (1.4) and (1.5) gives

$$S' = VS;$$
 $\frac{1}{f'} = \frac{1}{VS} - \frac{1}{S} = \frac{1-V}{VS};$ $S = f'\frac{1-V}{V};$ $S' = f'(1-V).$

Therefore

$$S' + S = \frac{f'(1 - V^2)}{V} = a;$$
 $S' - S = -f'\frac{(1 - V)^2}{V} = L$

(remember that V < 0, S < 0, and S' > 0 in both positions 1 and 2 of the lens). Solving the last equations for V and for f' we get

$$f' = \frac{aV}{1 - V^2} = -\frac{LV}{(1 - V)^2}; \quad V = -\frac{L + a}{L - a}; \quad f' = \frac{L^2 - a^2}{4L} = \frac{L}{4} - \frac{a^2}{4L}.$$

It is understood that linear magnifications V_1 and V_2 in positions 1 and 2 are reciprocal values ($V_1 = 1/V_2$), and the segments *S*, *S'* are just replacing each other while moving from position 1 to 2.

In deriving the above expressions we did not take into account the thickness of the lens, or, more exactly, the distance Δ between the principal planes. The rigorous relation is $L = S' - S + \Delta$. Actually Δ is unknown and therefore it is the origin of uncertainty in the value of *L*. Differentiating the above expression for f'with regard to *L* and denoting $dL = \Delta$ we obtain

$$\mathrm{d}f = \frac{\Delta}{4} \left(1 + \frac{a^2}{L^2} \right).$$

Now for the numerical data of the problem we have

$$f' = \frac{135}{4} - \frac{45^2}{4 \times 135} = 30.0 \text{ mm}.$$

If the thickness of the lens is about 6 mm then the distance between its principal planes is about 2 mm (approximately one-third of the lens thickness). Hence, for the uncertainty of the focal length we have

$$df = \frac{2}{4} \left[1 + \left(\frac{45}{135} \right)^2 \right] = 0.56 \text{ mm.}$$

1.5. Consider the layout of the thick lens shown in Fig. 1.35. We use Eqs. (1.7) and (1.8) and apply them to two surfaces of the lens. We choose an arbitrary value for h_1 and start with $u_1 = 0$, remembering that in our case $n_1 = n_3 = 1$; $n_2 = n$. Since we are looking for a solution in the paraxial range, where the heights of all rays are small, one can neglect the segments x_1 , x_2 (the latter is not shown in the



FIGURE 1.35 Problem 1.5 – Finding the location of the principal planes in a thick lens.

figure) assuming that the distance between h_1 and h_2 is equal to the lens thickness, t. Then we get

$$u_{2} = \frac{1}{n}u_{1} + \frac{h_{1}}{R_{1}n}(n-1) = \frac{h_{1}}{R_{1}n}(n-1); \quad h_{2} = h_{1} - u_{2}t = h_{1} - \frac{h_{1}t}{R_{1}n}(n-1);$$

$$u_{3} = nu_{2} + \frac{h_{2}}{R_{2}}(1-n) = \frac{h_{1}}{R_{1}}(n-1) - \frac{h_{1}}{R_{2}}(n-1) + \frac{h_{1}t}{R_{1}R_{2}n}(n-1)^{2}$$

which enables one to calculate the focal length and $S_{F'}$ (BFL):

$$\frac{1}{f'} = \frac{u_3}{h_1} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) + \frac{t(n-1)^2}{R_1R_2n}$$
$$S_{\mathrm{F}'} = \frac{h_2}{u_3} = f'\left[1 - \frac{t(n-1)}{R_1n}\right].$$

To find the segment S_F (FFL) we should repeat the same procedure, but to assume that the ray which is parallel to OX is incident on the surface R_2 of the lens from the right. Then the exit ray intersects the optical axis in the front focus F (left of the lens) and replacing f' by f and R_1 by R_2 in the above expression for $S_{F'}$ we finally get

$$S_{\rm F} = f \left[1 + \frac{t(n-1)}{R_2 n} \right].$$

Here we should remember that in our problem f < 0 and $R_2 < 0$, hence the value of S_F is negative. Using the numerical data of the problem we obtain

$$\Phi = \frac{1}{f'} = \left[0.5163\left(\frac{1}{50} + \frac{1}{75}\right) - \frac{6(0.5163)^2}{1.5163 \times 50 \times 70}\right] \times 10^3 = 16.92 \text{ dioptry}$$



FIGURE 1.36 Problem 1.6 – Ray tracing through a parallel plate.

$$f' = \frac{1}{\Phi} = 59.1 \text{ mm}; \quad S_{\mathrm{F}'} = 59.1 \left[1 - \frac{6 \times 0.5163}{50 \times 1.5163} \right] = 56.69 \text{ mm};$$
$$S_{\mathrm{F}} = -59.1 \left[1 - \frac{6 \times 0.5163}{75 \times 1.5163} \right] = -57.49 \text{ mm}.$$

As we see, the principal planes H and H' are located 1.61 mm and 2.41 mm, respectively, inside the lens.

1.6. Consider the ray incident on the slab at a height h_1 along the direction of the angle *i* (see Fig. 1.36). We have $h_2 = h_1 - t \times \tan(r)$ where the refraction angle *r* is calculated from Eq. (1.1). Since the incident angle at point 2 is also *r*, the refraction angle here (found again from Eq. (1.1)) is equal to *i*, meaning that the exit ray is parallel to the incident one. Then $O'A' = h_2/\tan(i) = h_1/\tan(i) - t \tan(r)/\tan(i)$; $OA = h_1/\tan(i)$; and therefore $AA' = O'A' - (OA - t) = t - t[\tan(r)/\tan(i)]$

$$AA' = t \left[1 - \sqrt{\frac{1 - \sin^2(i)}{n^2 - \sin^2(i)}} \right].$$

As we see, AA' depends on *i*, which means that each ray of the homocentric incident beam intersects the optical axis after the slab in another point A'. In other words, the homocentricity of the beam is violated. As the measure of this violation one can choose the value $\delta = (AA')_{i \text{ max}}$. Since i_{max} is related to the given solid angle, ω , as $\omega = 2\pi [1 - \cos(i_{\text{max}})]$, we obtain

$$\cos(i_{\max}) = 1 - \frac{\omega}{2\pi} = 1 - \frac{1.5}{2\pi} = 0.761; \quad \sin(i_{\max}) = 0.649$$
$$\delta = 5 \left[1 - \sqrt{\frac{0.761}{2.25 - 0.421}} \right] = 1.775 \text{ mm}; \quad O'A' = 31.775 \text{ mm}$$

[Note: The above expression for AA' is rigorous, it is valid for any angle *i*. For small angles *i* (paraxial approximation) we have $\sin i \approx i$; $\sin^2(i) \ll 1 < n^2$;



FIGURE 1.37 Problem 1.7 – Ray tracing through a system of two thin lenses.

 $AA' \approx t(1 - 1/n)$; and AA' does not depend on *i*. If n = 1.5 then AA' = t/3. This means that in the paraxial approximation the center of the incident beam is just relocated with regard to initial point A by one-third of the glass slab thickness (1.667 mm in our case).]

1.7. Considering a two-lens system in general, and referring to Fig. 1.37 one obtains

$$S_{2} = f_{1}' - d; \quad \frac{1}{S_{2}'} = \frac{1}{f_{2}'} + \frac{1}{S_{2}} = \frac{f_{1}' - d + f_{2}'}{f_{2}'(f_{1}' - d)}; \quad S_{2}' = \frac{1 - \Phi_{1}d}{\Phi_{1} + \Phi_{2} - \Phi_{1}\Phi_{2}d};$$
$$h_{2} = h_{1}\frac{S_{2}}{f_{1}'} = h_{1}(1 - \Phi_{1}d)$$

$$f'_{\rm e} = S'_2 \frac{h_1}{h_2} = \frac{1 - \Phi_1 d}{\Phi_1 + \Phi_2 - \Phi_1 \Phi_2 d} \times \frac{h_1}{h_1 (1 - \Phi_1 d)} = \frac{1}{\Phi_1 + \Phi_2 - \Phi_1 \Phi_2 d} = \frac{1}{\Phi_{\rm e}}.$$
 (A)

Now, by substituting the problem data in expression (A) we get

$$\Phi_{\rm e} = \frac{1}{100} + \frac{1}{75} - \frac{30}{100 \times 75} = 0.01933 \,\,{\rm mm^{-1}} = 19.33 \,\,{\rm diopter};$$

$$f'_{\rm e} = \frac{1}{0.01933} = 51.72 \,\,{\rm mm}; \quad S'_2 = \frac{1 - 0.01 \times 30}{0.01933} = 36.2 \,\,{\rm mm}.$$

Thus, the two lenses could be considered as a single system with 51.72 mm focal length and the focal point positioned 36.2 mm behind the second component. [Note: Replacing two lenses by a single equivalent lens is useful only if a parallel beam strikes the system. If imaging is performed for an object positioned at a final

distance from the first lens then Eq. (A) above becomes useless and calculations should be done according to Eqs. (1.4) and (1.5), first for the first element and then for the second one.]

1.8. For a ball lens of radius *r* the Eqs. (1.9) and (1.11) are transformed as follows $(d = 2r; r_1 = -r_2)$:

$$S_{\mathrm{F}'} = f' \left[1 - \frac{2(n-1)}{n} \right] = f' \frac{2-n}{n}; \quad \frac{1}{f'} = \frac{2(n-1)}{r} - \frac{2(n-1)^2}{rn} = \frac{2(n-1)}{nr}$$
(A)

and therefore

$$a' = f' - S_{F'} = f'\left(1 - \frac{2-n}{n}\right) = \frac{nr}{2(n-1)} \times \frac{2(n-1)}{n} = r.$$
 (B)

Thus, the principal plane H' is located at the center of the ball. Due to the symmetry of the lens one can state that the front principal plane is located at the same point. From the data of the problem, using the glass data from Appendix 2 ($n_D = 1.67270$), we find

$$S_{\rm F} = \frac{2 - 1.6727}{2 \times 0.6727} 3 = 0.73 \text{ mm}; \quad f' = \frac{3 \times 1.6727}{2 \times 0.6727} = 3.73.$$

As we see, the focus is distant from the lens surface by 0.73 mm.

1.9. The angle in air between two chief rays directed to two separate object points still distinguished by the eye is 3×10^{-4} rad. Taking into account the "reduced eye" properties, in particular the refractive index of the medium between the eye lens and the retina as n = 1.336 and the back focal length as 22.9 mm, we get that the limiting angle in the vitreous is $3 \times 10^{-4}/1.336 = 2.25 \times 10^{-4}$ rad. The corresponding distance between two images on the retina is $2.25 \times 10^{-4} \times 22.9 = 5.15 \times 10^{-3}$ mm and they should fall on two different cells. This means that the retina cell size is about 5µm.

1.10. (a) The intermediate image in the branch to the eye is formed in the plane of the reticle M of size 19 mm. As linear magnification of the objective is $V_1 = -10$, it yields

$$S'_1 = -10S_1;$$
 $\frac{1}{-10S_1} - \frac{1}{S_1} = \frac{1}{20};$ $S_1 = -22 \text{ mm}$

and this is the working distance of the system. The corresponding field of view is 1.9 mm.

(b) The eyepiece has visual magnification determined from Eq. (1.15): $\Gamma = 250/25 = 10$ and therefore the total magnification in the branch to the eye is $V_{\text{tot}} = V_1 \Gamma = 100$.



FIGURE 1.38 Problem 1.10 – Formation of an image onto the CCD plane.



FIGURE 1.39 Problem 1.10 - Formation of image on CCD, the second case.

(c) The optimal imaging in the CCD branch is the one which enables one to see the maximum part of the image created on the circular reticle M. Such a situation shown in Fig. 1.38 means that the maximum image size on the CCD is $y' = \sqrt{(4.8^2 + 5.6^2)} = 7.4$ mm and therefore the relay lens L₃ provides an optical magnification $V_3 = 7.4/19 = 0.388$. This can be realized in two possible arrangements. The first is demonstrated in Fig. 1.38 and the second in Fig. 1.39. In both cases $S'_3 = 20$ mm, but $S_3 = S'_3/V_3 = 20/0.388 = 51.5$ mm in the first case and $S_3 = -51.5$ mm in the second case. Evidently the shortest configuration is that of Fig. 1.38. In this case

$$\frac{1}{f'_3} = \frac{1}{20} - \frac{1}{51.5}; \quad f'_3 = 32.7 \text{ mm}$$

and the distance between P and the CCD is $22 \times 10 - 51.5 = 168.5$ mm. In the second case

$$\frac{1}{f'_3} = \frac{1}{20} + \frac{1}{51.5}; \quad f'_3 = 14.4 \text{ mm}$$

and the distance between P and the CCD is 220 + 51.5 = 271.5 mm.

1.11. (a) From the numerical data of Fig. 1.22 we have $S_1 = -70$; $S'_1 = (-70) \times (-3) = 210$. Therefore, the distance between the CCD and the last beam splitter

is 210 - (10 + 100 + 70) = 30 mm and the focal length of L₁ should be

$$f_1' = \left(\frac{1}{210} + \frac{1}{70}\right)^{-1} = 52.5 \text{ mm.}$$

In the high-magnification branch the image created by lens L₁ (the same size and position as in the low-magnification branch) serves as the virtual object for the second lens, L₂ (negative relay lens). Magnification of L₂ is $V_2 = 2 \times (-3)/V_1 = 2$. Since the distance along the optical axis between L₁ and the CCD is $l = 10 + 70 + 2 \times (10 + 15) + 100 + 30 = 260$ mm and taking into account that $S'_2 - S_2 = l - S'_1 = 50$ mm and $S'_2 = 2 \times S_2$, we get $S_2 = 50$ mm and $S'_2 = 100$ mm and therefore the location of L₂ is 70 mm below the last beam splitter. Its focal length is

$$f_2' = \left(\frac{1}{100} - \frac{1}{50}\right)^{-1} = -100 \text{ mm}.$$

(b) Increasing the high magnification by 10% requires $V_2 = 2.2$ (V_1 remains the same as before). Hence,

$$S'_2 = 2.2S_2; \quad \frac{1 - V_2}{V_2 S_2} = \frac{1}{f'_2};$$

which yields $S_2 = 54.54$ mm and $S'_2 = 120$ mm. In other words, the relay lens should be relocated to 90 mm below the last beam splitter. Since in this case $S'_2 - S_2 = 120 - 54.54 = 65.46$ mm it is necessary to add the length 15.46 mm to the optical path of the left branch. This is done by moving down the retroreflector by the segment $\Delta z = 15.46/2 = 7.73$ mm.

1.12. (a) To get an image with no vignetting it is necessary to position the field stop in the plane of the intermediate image created by the first lens. Referring to Fig. 1.40 we get

$$S_1 = -200;$$
 $S'_1 = \left(\frac{1}{f'_1} + \frac{1}{S_1}\right)^{-1} = \left(\frac{1}{100} - \frac{1}{200}\right)^{-1} = 200 \text{ mm};$ $V_1 = -1;$

and therefore the field stop should be of the same size as the field of view, i.e., $d_{ab} = 10$ mm, and positioned 200 mm behind L₁. The total magnification $V = 3 = V_1 \times V_2$ requires $V_2 = -3$, which enables one to find the position of L₂ and M:

$$\frac{1-V_2}{S_2V_2} = \frac{1}{f_2'}; \quad S_2 = 50\frac{1+3}{(-3)} = -66.67 \text{ mm}; \quad S_2' = S_2V_2 = 200 \text{ mm}.$$

(b) The field stop size, as we saw above is 10 mm.



FIGURE 1.40 Problem 1.12 – Imaging system with the field stop.

(c) To find the entrance pupil we should build the image of all diaphragms (L_2 and ab in our case) in the object space, e.g., to create their images through L_1 at reverse illumination (as if radiation propagates from right to left). Since ab is conjugated with the object plane P, one has to find only the image of L_2 through the first lens. We have:

$$S_{21} = -266.67 \text{ mm};$$
 $S'_{21} = \left(\frac{1}{100} - \frac{1}{266.67}\right)^{-1} = 160.0 \text{ mm};$
 $d'_{L_2} = 20 \times \frac{160}{266.67} = 12 \text{ mm}.$

Calculating the angle of the margin ray coming from the on-axis point of the object to the side point of the lens L_1 gives $\alpha_1 = 10/200 = 0.05$. The corresponding angle of the image of L_2 is $\alpha_2 = 6/(200 - 160) = 0.15 > \alpha_1$ and therefore the entrance pupil is the lens L_1 and the aperture angle of the system is $\alpha_{ap} = \alpha_1 = 0.05$.

1.13. Referring to the solution of Problem 1.10 and Fig. 1.41, we first find the size of the lens L_1 using NA = 0.2 and the distance to the object $S_1 = -22$ mm: $D_{L1} = 2 \times 22 \times \tan(\arcsin 0.2) = 9.0$ mm. Then we proceed with the margin



FIGURE 1.41 Problem 1.13 – The margin ray tracing through lenses L_1 and L_3 .



FIGURE 1.42 Problem 1.14 – Two approaches to finding the image: (a) without unfolded diagram; (b) with unfolding.

ray originating in the off-axis point of object A. This ray comes to the side point A' of the intermediate image (O'A' = 19/2 = 9.5 mm) and it is this ray which determines the active size of lens L₃. Geometrical consideration of the figure gives

$$D_{L3} = 2 \times (O'_1 N + ND) = 2 \times \left[D_{L1}/2 + \frac{168.5}{220} (9.5 - 4.5) \right] = 16.66 \text{ mm}.$$

1.14. To demonstrate the advantage of the unfolded diagram we describe two approaches in solving the problem: first without the diagram and then using unfolding. In the first case we start with imaging through mirror M_1 (see Fig. 1.42a) and find the image point A' using the triangle AO_1A' , where $AO_1 = (80 - 20) = 60 \text{ mm} = A'O_1$. Obviously the second image point, B', is on the horizontal line passing through A'. Then, referring to A'B' as a new object we find its image in mirror M_2 : $A''O_2 = A'O_2 = 80 \text{ mm}$ and B'' is again located on the horizontal line passing through A''.

In the second case (Fig. 1.42b) we create the image of the mirror corner in mirror M_1 . Then mirror M_2 image, M'_2 , is vertical and A'B' is parallel to AB and distant from M'_2 by 80 mm. Going back to the real mirror M_2 we just put A''B'' beneath M_2 at the same distance 80 mm and 20 mm to the right of the vertex. As we see, the second approach is significantly shorter and easier.

1.15. We build an unfolded diagram for the prism, as demonstrated in Fig. 1.43, and consider the principal ray MNPQ striking the entrance face AB at the height $AM = (a \times \sin 45^\circ)/2 = 0.707$ cm. This is the center of a beam passing through the prism. Obviously $MN = AM = PQ = P_1Q_1 = 0.707$ cm, $NP_1 = a = 2$ cm, and $t_e = MN + NP_1 + P_1Q_1 = 3.414$ cm. Hence, the apparent (reduced) thickness is

$$\frac{t_{\rm e}}{n} = \frac{3.414}{1.5163} = 2.251 \,{\rm cm}.$$



FIGURE 1.43 Problem 1.15 – Unfolded diagram of a rhomboidal prism.



FIGURE 1.44 Problem 1.16 – Imaging through a penta-prism.

1.16. We refer to Fig. 1.44 and assume that the lens is working in the paraxial range. Without the prism the distance from the lens to the screen P would be

$$S' = \left(\frac{1}{30} - \frac{1}{40}\right)^{-1} = 120 \text{ mm.}$$

The thickness of the glass block which is equivalent to the prism is $t_e = 3.41a = 34.1$ mm. The prism makes the ray trajectory longer by the segment $\Delta = t_e(1 - 1/n) = 34.1(1 - 1/1.5163) = 11.61$ mm. Finally, from the geometry of the figure we get for the distance between the screen P and the exit face of the prism x = 120 - 20 - 34.1 + 11.61 = 88.39 mm.

1.17. The prism ABC of refractive index *n* has a vertex angle β and an input ray strikes the side AB at point O₁ at an incident angle *i*₁ (see Fig. 1.45). The deviation angle φ is defined as the angle between the input direction and the output direction of the ray. Geometrical consideration of the triangles O₁BO₂ and O₁DO₂ yields

$$\varphi = (i_1 - r_1) + (r_2 - i_2); \quad r_1 + i_2 + \gamma = 180^\circ = \beta + \gamma$$

$$r_1 + i_2 = \beta$$
(A)

$$\varphi = i_1 + r_2 - \beta. \tag{B}$$



FIGURE 1.45 Problem 1.17 – Deviation of a ray traveling through a prism ABC.

Snell's law gives $i_1 = \arcsin(n \sin r_1)$ and $r_2 = \arcsin(n \sin i_2) = \arcsin[n \sin(\beta - r_1)]$. By substituting these expressions in (B) we get

$$\varphi = \operatorname{arc}(n \sin r_1) + \operatorname{arcsin}[n \sin(\beta - r_1)] - \beta.$$
 (C)

To find the minimum deviation angle we calculate the derivative $d\varphi/dr_1$ and find the angle at which it has a zero value, as usual:

$$\frac{d\varphi}{dr_1} = \frac{n\cos r_1}{\sqrt{1 - n^2\sin^2 r_1}} - \frac{n\cos(\beta - r_1)}{\sqrt{1 - n^2\sin^2(\beta - r_1)}} = 0;$$
$$\cos r_1 \sqrt{1 - n^2\sin^2\psi} = -\cos\psi\sqrt{1 - n^2\sin^2 r_1} = 0$$

where the new variable $\psi = \beta - r_1$ is introduced. From the last equation we have

$$\frac{\cos^2 r_1}{\cos^2 \psi} = \frac{1 - n^2 \sin^2 r_1}{1 - n^2 \sin^2 \psi};$$

and denoting $z = \sin^2 r_1$ we proceed as follows:

$$\frac{1-z}{\cos^2\psi} = \frac{1-n^2z}{1-n^2\sin^2\psi};$$

$$z = \frac{1/\cos^2\psi - 1/(1-n^2\sin^2\psi)}{1/\cos^2\psi - n^2/(1-n^2\sin^2\psi)}$$

$$= \frac{\sin^2\psi \times (1-n^2)}{1-n^2} = \sin^2\psi.$$

The last equation is satisfied if $r_1 = \psi$ and therefore $r_1 = \beta - r_1$; and $r_1 = \beta/2$. With this value we have from (C):

$$\varphi = 2 \arcsin\left(n \sin\frac{\beta}{2}\right) - \beta$$
 (D)

and

$$i_1 = \arcsin\left(n\sin\frac{\beta}{2}\right).$$
 (E)

The last two expressions allow one to calculate the angle of minimum deviation of the prism and the incidence angle corresponding to such a deviation. Going back to the problem, we find $\varphi = 2 \arcsin(1.6727 \times \sin 30^\circ) - 60^\circ = 53.51^\circ$ and the incidence angle $i_1 = \arcsin(1.6727 \times \sin 30^\circ) = 56.76^\circ$.

This page intentionally left blank

Theory of Imaging

2.1. Optical Aberrations

2.1.1. General Consideration. Ray Fan and Aberration Plot. Concept of Wave Aberrations

We will proceed by considering the concept of imaging as described in Section 1.2 of Chapter 1 and pay most attention to the real imaging situation experienced in practice. Figure 2.1 demonstrates the basic difference between ideal imaging and real imaging. Let the rays originating in a point source A come to the system, each one at a different angle u. If the medium is homogeneous (has the same refractive index everywhere) the wavefront W in the object space is a sphere. If in the image space *all rays* intersect at a single point A' then the beam remains homocentric, with a spherical wavefront W', and A' is a stigmatic (ideal) image of A. However, in most situations this does not happen and the rays of different angles u' come to different points on the axis OO' (or, for tilted beams, to different off-axis locations). As a result, the real wavefront in the image space is not spherical, the homocentricity of the output beam is violated, and instead of a sharp point image there is a blurred spot. Such violation of stigmatic imaging is defined as optical aberrations.

Numerically aberrations are characterized by the deviation of a real image A' from the ideal image A'_0 obtained in the paraxial range. This deviation can be determined either by the horizontal segment, $\delta s'$, along the optical axis, as in Fig. 2.2 (and then it is called the lateral aberration) or it can be related to the vertical segment ρ (then it is called the transverse aberration). The geometrical



FIGURE 2.1 (a) Ideal imaging and (b) real imaging.



FIGURE 2.2 (a) Lateral and transverse aberration and (b) the aberration diagram.

relation between lateral and transverse aberrations is quite obvious:

$$\rho = \delta s' \times \tan u \approx \delta s' \frac{h}{S'} \tag{2.1}$$

in which the fact is taken into account that $\delta s' \ll S'$. Since for each ray aberrations depend on the height of the ray on the last refractive surface, and consequently on the whole optical path while it travels through the optical system, it is commonly accepted to represent the aberrations by a diagram like that shown in Fig. 2.2b. The graph always passes through the zero point, meaning that at very small heights, e.g., in paraxial range, there are no aberrations (there is an exception to this rule, which is considered in Section 2.2).

There is a great variety of reasons why aberrations happen in optical systems. Some of them are relevant in a specific application whereas some others are not. It was understood at a very early stage of the development of aberration theory that it is worth classifying aberrations in three groups and consider each one separately. These groups are:

- (a) chromatic aberrations chromaticity of location (the only aberration existing also in the paraxial area) and chromaticity of magnification;
- (b) monochromatic aberrations of wide beams (spherical aberration and coma);

(c) field aberrations or monochromatic aberrations of tilted beams (astigmatism, field curvature, and distortion).

We will address each group in following sections of this chapter.

To characterize the image quality it is not enough to consider aberrations of several rays coming from an on-axis point. It is necessary to analyze a great number of rays coming from on-axis as well as from off-axis points of the object and to do this for three wavelengths at least if the system is intended for imaging with white light. Usually the ray tracing analysis is carried out for rays propagating in the vertical plane passing through the optical axis (this plane is called the tangential or meridional plane) and for rays propagating in a tilted plane where an off-axis point of the object and horizontal diameter of the entrance pupil are located (this is called the saggital plane). More specifically (see Fig. 2.3), a number of points on the vertical and horizontal diameters of the entrance pupil are chosen and the meridional fan of rays (all in the plane TP) and the saggital fan of rays (all in the tilted plane SP) are analyzed aiming at the location of the final destination of each ray in the image plane. Then the meridional plot and the saggital plot, like the two graphs shown in Fig. 2.3b, are created followed by the calculation, if necessary, of some other integral parameters of the image (like spot diagrams, energy distribution, vignetting rate, modulation transfer function, etc.).

With regard to the integral characteristics of imaging, one more issue should be considered here. Image blurring can be characterized not only in terms of the geometric parameters of the rays but also in terms of wavefront distortion or wave aberrations (also termed optical path differences, OPDs). Referring to Fig. 2.1b, consider the real wavefront W' and the virtual reference sphere (dotted line) of radius S' centered at the point A'_0 . The distance *l* between W' and the reference sphere along the radius passing through A'_0 and tilted to the axis at an angle u''



FIGURE 2.3 (a) The fan of rays in the entrance pupil and (b) the meridional and saggital plots.

is called the wave aberration, or OPD. The difference between the angles u' and u'' is small, so that the local wave aberration in terms of lateral aberration can be expressed as

$$l = n \times \delta s' \times (1 - \cos u') \tag{2.2}$$

and the overall (cumulated) wave aberration is defined by the integral

$$l = n \int_{0}^{u} \delta s' \times \sin u' \mathrm{d}u'. \tag{2.3}$$

The OPD value can be calculated if aberrations $\delta s'$ are known for all angles from 0 to *u*. The expression for lateral aberrations in terms of wave aberration is of great importance and allows one to obtain analytical expressions for lateral and transverse ray aberrations in a closed form as far as a third-order approximation is considered (Seidel's formula, discussed in following sections of this chapter).

There exists an important Rayleigh's criterion of acceptable degradation due to aberrations: the image quality is acceptable if the wave aberration l is not greater than 0. 25 λ .

Problems

2.1. A lens L of 10 mm diameter and 100 mm focal length working in the paraxial range builds a sharp image at magnification V = -2 in the plane P where the observation screen is located. If L is replaced by another lens of the same nominal focus but manufactured with 5% tolerance, what blurring could be expected on the screen?

[Note: Calculate the meridional plot of rays in the plane P.]

2.2. A lens of 40 mm size designed to form an image at a distance of 125 mm in air was used in a laboratory set-up where the optical axis was turned through 90° by a penta-prism of 30 mm entrance face positioned 35 mm behind the lens. Assuming the prism is made of BK-7 glass (n = 1.5163) find the meridional ray plot of the additional aberration introduced by the prism.

2.1.2. Chromatic Aberrations: Principles of Achromatic Lens Design

As we mentioned earlier, chromaticity caused by chromatic aberration of location is the only aberration (except defocusing) experienced even in the paraxial range. The origin of chromaticity is in the dispersion of light inside optical elements (made of glass or crystals). It is well known that the refractive index of optical glasses varies with wavelength and its spectral behavior can be approximately described by the formula

$$n(\lambda) = A + \frac{B}{(C - \lambda)^2}$$
(2.4)

where *A*, *B*, and *C* are constants characterizing a specific material. Usually the refractive index is considered for three main wavelengths, $\lambda_D = 0.589 \,\mu\text{m}$; $\lambda_F = 0.486 \,\mu\text{m}$, and $\lambda_C = 0.656 \,\mu\text{m}$, and the corresponding values n_D , n_F , and n_C are also included in the parameter of dispersion called the Abbe number (or the Abbe value):

$$v_{\rm D} = \frac{n_{\rm D} - 1}{n_{\rm F} - n_{\rm C}}.$$
 (2.5)

Selected data for several optical glasses are presented in Appendix 2.

Since the focal length of a lens is directly related to its refractive index by Eq. (1.11), it is quite understandable that if the lens is operated simultaneously at several wavelengths (or with white light illumination) significant chromatic aberration might occur in the system. Usually chromatic aberration is defined as the difference between the focal length at the selected wavelength relative to that of line D:

$$\delta_{\rm Ch} = f_{\lambda}' - f_{\rm D}'. \tag{2.6}$$

In more general cases δ_{Ch} is related to the distances between the lens and the image, $\delta_{Ch} = S'_{\lambda} - S'_{D}$, and apparently varies with the magnification of the system.

Chromatic aberration of a single lens is demonstrated in Fig. 2.4 (curve 1) and explained by the ray diagrams of Fig. 2.5 separately for positive and negative lenses. As can be seen, the aberration plots in these two cases are opposite and this fact is widely exploited for the correction of chromaticity. The lens is divided in two components, one positive and one negative, which are designed according to the rules described below and then brought in contact and cemented in a single element called a doublet lens, or achromat.

From the variety of available optical glasses we choose two different materials – one for the positive component (Abbe value v_{D1}) and another for the negative part (with Abbe value v_{D2}). Neglecting the thickness of the components we have for the total optical power, Φ , of the achromat

$$\Phi = \Phi_1 + \Phi_2. \tag{2.7}$$



FIGURE 2.5 Chromatic aberration of (a) positive and (b) negative lenses.

Each component obeys the single lens equation (1.11) which we rewrite in the form

$$\Phi_k = (n-1)c_k \tag{2.8}$$

where $c_k = (1/r_{k1} - 1/r_{k2})$ is the bending parameter independent of wavelength. Considering the variation of the optical power as the wavelength is changed from λ_F to λ_C , we have

$$\mathrm{d}\Phi_k = \frac{\Phi_k}{v_{\mathrm{D}k}}.\tag{2.9}$$

Although $d\Phi_1$ and $d\Phi_2$ both have finite values, we require that the variation of the total optical power be zero: $d\Phi = d\Phi_1 + d\Phi_2 = 0$, which yields the following equation:

$$\frac{\Phi_1}{\nu_{\rm D1}} = -\frac{\Phi_2}{\nu_{\rm D2}}.$$
 (2.10)

Resolving Eq. (2.10) together with Eq. (2.7) gives the following formula for the optical power of both components of the achromat:

$$\Phi_1 = \frac{v_{\text{D1}}}{v_{\text{D1}} - v_{\text{D2}}} \Phi; \quad \Phi_2 = -\frac{v_{\text{D2}}}{v_{\text{D1}} - v_{\text{D2}}} \Phi.$$
(2.11)

To complete the design of the achromat we have to find the bending parameters. Since we have only two conditions (2.11) for four independent radii, r_1 , r_2 , r_3 , and r_4 , there are two degrees of freedom here. One degree can be reduced if we require that the contacting surfaces of both lenses have the same shape, e.g., $r_2 = r_3$. An additional degree of freedom is one of the two remaining radii. Indeed, we can arbitrarily choose one of them (e.g., $r_4 = \infty$) and then complete the design in the following manner:

$$r_3 = \frac{n_{\rm D2} - 1}{\Phi_2} = r_2; \quad r_1 = \frac{n_{\rm D1} - 1}{\Phi_1}.$$

Or, choose the first surface of the positive lens to be plane and then calculate the rest of the shapes. Several possible forms of doublet lens are shown in Fig. 2.6. All of them, however, represent a cemented pair (the adhesive used is of a refractive index very close to that of glass).

There also exists the possibility of designing an achromatic lens with an air spacing between the components (e.g., see the detailed explanation in Kingslake, 1979). In any case the achromatic lens provides two focuses to coincide, F_F and F_C . The remaining difference between these two and the focus of the line D is called the secondary spectrum (it is shown by curve 2 in Fig. 2.4). In some situations the residual chromatic aberration of the doublet lens is too large and further correction is required. This is realized in triplet lenses or in more complex configurations. The remaining chromatism is called the tertiary spectrum (curve 3 in Fig. 2.4). It can be seen that the three focuses coincide in such a case and the residual aberration is very small.



FIGURE 2.6 Doublet lenses of different shapes.

Problems

2.3. Find the chromatic aberration introduced by the penta-prism in Problem 2.2 and build the aberration plot.

2.4. Find the doublet lens of 13.33 diopters optical power if the components are made from BK-7 and F-1 glasses and calculate the residual chromatic aberration (the secondary spectrum).

2.1.3. Spherical Aberration and Coma

These two types of aberrations are monochromatic aberrations of a wide beam. Consider first the spherical aberration (see Fig. 2.7). Due to the geometry of a spherical shape the rays originating in an on-axis point A and incident on the lens at different distances *h* from the optical axis are not concentrated in a single point behind the lens, but create images at separate locations $(A'_1, A'_2, A'_3, \text{ etc.})$. The lateral spherical aberration $\delta s'_{\text{Sph}}(h_i)$ is defined as the distance between the image in the paraxial range (A'_1) and the image corresponding to the height h_i (e.g., the point A'_i). The corresponding transverse spherical aberration $\delta s'_{\text{Sph}}$ defines the size of the light spot created in the plane perpendicular to the axis (e.g., on an observation screen). At any position along the axis the spot on the screen has a finite size, but at some location the size is a minimum and this is the point of the best imaging, as far as spherical aberration is concerned.



FIGURE 2.7 Spherical aberration of (a) a single positive lens and (b) a single negative lens.

To describe the spherical aberration analytically it is usually expanded in a power series

$$\delta s'_{\text{Sph}}(h) = a_3 h^2 + a_5 h^4 + \dots$$
 (2.12)

(it can be easily shown that the terms with coefficients a_0 and a_1 are equal to zero). If only the first term of Eq. (2.12) is considered then the solution can be derived in a closed form. Such an approximation is called a third-order aberration and it is commonly known as Seidel's formula. We describe it here as follows (for further details, see Born and Wolf, 1968):

$$\delta s'_{\text{Sph}} = -\frac{1}{2} \frac{h^2 S'^2}{(1-\xi)^2 f'} \left[A + B \frac{1-\xi}{r_1} + (1+2\xi) \left(\frac{1-\xi}{r_1}\right)^2 \right]$$
(2.13)

where

$$A = \frac{3}{S'}C_1 + \frac{1}{f'^2} - \xi(2+\xi)C_2C_1 + \xi^2(1+2\xi)C_1^2$$

$$B = 2\xi(1+2\xi)C_1 - (2+\xi)C_2$$

$$C_1 = \frac{1}{S'} - \frac{1}{f'}; \quad C_2 = \frac{2}{S'} - \frac{1}{f'}; \quad \xi = \frac{1}{n}.$$

Expression (2.13) can be used to estimate the spherical aberration at any position of the object and the image. For the special case of the object in infinity, S' = f' and Eq. (2.13) is transformed to

$$\delta s'_{\rm Sph} = -\frac{1}{2} \frac{h^2 f'}{(1-\xi)^2} \left[\frac{1}{f'^2} - \frac{(2+\xi)}{f'} \frac{(1-\xi)}{r_1} + (1+2\xi) \left(\frac{1-\xi}{r_1}\right)^2 \right]. \quad (2.14)$$

As the aberration value depends explicitly on the shape of the lens (radius r_1), one might minimize aberration by optimizing the shape. $\delta s'_{Sph}$ achieves its minimum value

$$(\delta s'_{\text{Sph}})_{\min} = -\frac{1}{8} \frac{\xi}{(1-\xi)^2} \frac{(4-\xi)}{(1+2\xi)} \frac{h^2}{f'}$$
(2.15)

when its radii obey the relations:

$$r_1 = 2(1-\xi)\frac{1+2\xi}{2+\xi}f'; \quad r_2 = 2\frac{(1-\xi)(1+2\xi)}{2-\xi-4\xi^2}f'.$$
 (2.16)

For instance, assuming n = 1.5 and keeping in mind that $h_{\text{max}} = D/2$ we get from Eqs. (2.15) and (2.16)

$$\delta s'_{\text{Sph}} = -0.268 \frac{D^2}{f'}; \quad r_1 = \frac{7}{12} f'; \quad r_2 = -3.5 f'.$$
 (2.17)

One should remember that in the aberration blur the radiation energy is not equally distributed. For this reason half the size of the maximum spot caused by aberration and calculated from Eqs. (2.13)–(2.17) is exploited as an aberration measure. It should also be mentioned again that the above formulas enable one to estimate the spherical aberration of a single lens approximately. To get more rigorous results the ray tracing procedure is inevitably required.

As can be seen from Fig. 2.7, the lateral spherical aberrations of a positive lens are negative whereas the aberrations of a negative lens have the opposite sign. This fact allows one to reduce drastically the spherical aberration if the single lens is replaced by a doublet (like the achromat described in Section 2.1.2).

Coma is an aberration of a wide tilted beam originating in an off-axis point of the object. This aberration is caused by the fact that the magnification of the system is not constant, but varies with the height of the incident ray: V = F(h).

Figure 2.8 demonstrates the formation of coma and explains the parameter, δk , chosen as its numerical measure:

$$\delta k = \frac{1}{2}(y'_1 + y'_2) - y'_C \tag{2.18}$$

where y'_{C} is the vertical coordinate of the chief ray of the beam at the image plane P and y'_{1} and y'_{2} are the vertical coordinates of the upper and the lower rays 1 and 2 on the same plane. The ray bundle starting in the off-axis object point A and coming to the entrance pupil is not symmetrical with regard to the optical axis, so it is not surprising that the spot in the plane P is also not symmetrical. The conditions and methods of coma correction are discussed later in this chapter.



FIGURE 2.8 Formation of coma.

Problems

2.5. (a) Find the optimal shape of a lens of 30 mm diameter and $f^{\#} = 2.0$ (*f*-number, *f*#, defined as the ratio f'/D of a lens focus to its diameter) intended for imaging from infinity if it is made of (i) BK-7 glass and (ii) SF-11 glass, and calculate the maximum transverse aberration in both cases (for imaging in monochromatic light of wavelength *D*). (b) How will the results be changed if the lens is turned by 180° ?

2.6. Spherical aberration of a cylinder rod or a sphere: a rigorous ray tracing. (a) Calculate the plot of transverse spherical aberration of a cylinder rod of 7 mm diameter made of BK-7 glass working with a point light source (laser diode of 0.59 wavelength) located 2 mm in front of the rod. (b) How will the results of the calculation be affected if the rod is replaced by a lens having the shape of a full sphere of 7 mm diameter (a ball lens)?

2.7. Spherical aberration of a plano-convex cylindrical lens: a rigorous ray tracing. How will the plot of spherical aberration calculated in Problem 2.6 be changed if the cylinder rod is replaced by a plano-convex cylindrical lens of the same radius (3.5 mm) made of BK-7 glass? The plane P remains at the same location as in Problem 2.6. The size of the new lens is shown in Fig. 2.9.



FIGURE 2.9 Problem 2.7 – Plano-convex cylindrical lens and the image plane.

2.1.4. Aberrations of Tilted Beams (Field Aberrations)

This group of aberrations includes astigmatism, curvature of field, and distortion.

Astigmatism

This aberration occurs if a pencil of tilted rays originating in an off-axis point of the object strikes the entrance pupil of the system. Astigmatism is illustrated in Fig. 2.10. For a tilted beam (which is initially homocentric) the optical axis is not



FIGURE 2.10 Astigmatism of a single lens: (a) imaging by meridional and saggital rays; (b) cross-section of the light spots along the optical axis.

an axis of symmetry any more and the behavior of the rays in the meridional plane (rays 1 and 2) differs from that of the saggital rays (rays 3 and 4). As a result the lens concentrates the tangential rays and the saggital rays in two different points, A'_t and A'_s . Both are out of the plane P of the paraxial image (point A'_0). Aberration of astigmatism is measured as the distance between the meridional and saggital images originating in the same point of the object (in Fig. 2.10a $\delta s'_{Ast} = S'_t - S'_s$). Obviously the greater the height of point A the larger the difference $S'_t - S'_s$, and for the on-axis point O aberration of astigmatism is approaching zero. The cross-section of the light bundle behind the lens is not homocentric anywhere but varies in a manner demonstrated in Fig. 2.10b.

Astigmatic aberration appears not only in elements with optical power (lenses or mirrors), but features also in a parallel plate. In this case the aberration can be described analytically. Referring to Fig. 2.11, we consider the tangential and saggital images A'_t and A'_s of a point A having a (virtual) image A'_0 (e.g., the image that would be created in air, without a parallel plate of thickness *d*). The distance *a* between the two images is determined by the formula

$$a = \left(1 - \frac{\cos^2 u}{\cos^2 u'}\right) \frac{d}{n \cos u'} \tag{2.19}$$



FIGURE 2.11 Astigmatism in a parallel plate.

and the astigmatic aberration becomes

$$\delta s'_{\text{Ast}} = \delta = (n^2 - 1) \frac{\tan^3 u'}{\tan u} d \approx \frac{(n^2 - 1)}{n^3} u^2 d.$$
(2.20)

Curvature of Field

Going back to the astigmatism of a lens-based system as shown in Fig. 2.10, one may note the fact that both the tangential and saggital images are not segments of straight lines but rather have noticeable curvature. Furthermore, it is reasonable to assume that the image created simultaneously by tangential as well as by saggital rays is located on a curved surface passing somewhere between the meridional and saggital images, as depicted in Fig. 2.12. This is commonly defined as an



FIGURE 2.12 Occurrence of curvature of field.

additional aberration called the curvature of field and is estimated as the radius of curvature, ρ , of the best image. It can be shown that the value of ρ obeys the following expression (Petzval's theorem):

$$\frac{1}{\rho} = -n' \sum_{i} \frac{1}{r_i} \left(\frac{1}{n_i} - \frac{1}{n_{i-1}} \right)$$
(2.21)

where n' is the refractive index in the image space and the summation is carried out over all refraction surfaces of the system.

Distortion

It is assumed in paraxial optics that linear magnification between two conjugate planes is defined solely by the location of the planes along the optical axis (in other words, by the distance of the object to the lens). In reality this assumption is violated and linear magnification, V, depends not only on the location of the plane along OZ, but also on the distance of the point of interest from the optical axis (distance in the radial direction). Violation of the above condition causes distortion of images, as illustrated in Fig. 2.13. The object shown is a regular square of size a with its center O positioned on the optical axis. Since the radial distances from O to points A and B are different, a/2 and $a/\sqrt{2}$, respectively, their images A' and B' are determined by different magnifications, V_A and V_B , and the whole image of the square is deformed. Distortion is characterized numerically as follows:

$$\Delta = \frac{y' - y'_0}{y'_0} 100\% \tag{2.22}$$



FIGURE 2.13 (a) Distortion and (b) positive and negative distortion of images.



FIGURE 2.14 Distortion in a parallel plate.

where y'_0 and y' are the radial displacement (or height) of the paraxial image and the real image of the same point. The value defined by Eq. (2.22) is sometimes called the fractional distortion.

Two kinds of distortion can be experienced in imaging systems, positive and negative. In the first case linear magnification in the image plane is increased with radial distance. In the second case the larger the distance the lower the magnification. Both cases are shown in Fig. 2.13b.

Distortion might originate not only in lenses or mirrors, but also in prisms or parallel glass plates. Such a case is shown in Fig. 2.14 where the off-axis image, y', created by the system is transferred by the parallel plate of thickness d into the final image y''. Distortion can be expressed in terms of the thickness and refractive index of the plate and the skew angle u and the distance p from the entrance pupil to the image plane:

$$\Delta = -\frac{n^2 - 1}{2n^3} \frac{d}{p} u^2.$$
 (2.23)

Aberration of distortion might be very critical in some applications, for example in optical systems for mapping.

Problems

2.8. A lens of 30 mm focal length operates in an angular field of view of $\pm 30^{\circ}$ and creates an image at magnification V = -2. Behind the lens, at a distance of 20 mm from it, a right-angle prism of 30 mm \times 30 mm size is positioned in order to bend the optical axis by 90°. Find the diagram of astigmatism and distortion across the field of view.

2.9. A flattener element in the imaging system. A bi-convex symmetrical lens of 40 mm focal length made of BK-7 glass performs imaging of distant objects to the plane P in a wide field of view. Is it reasonable to expect that the image quality of the off-axis areas will be of the same grade as images close to the optical axis? Find the flattener which is capable of improving image degradation for off-axis points (assume that it is made of SF-11 glass).

2.1.5. Sine Condition and Aplanatic Points

Once we realize that imaging in general is accompanied by aberrations, it is quite understandable that finding locations where aberrations are small or even can be avoided completely is of great importance not only from a theoretical point of view but also for practical reasons. It can be shown that such locations do exist for a single surface with curvature, either a reflective or refractive surface, aspherical or spherical. Obviously the latter is more attractive, since manufacturing spherical optics is much easier and cheaper than fabrication of aspherical elements.

Let us consider a refraction surface Q separating media of refractive index *n* and *n'*, as illustrated in Fig. 2.15, and let the conjugate pair A and A' be the points of stigmatic imaging (imaging with homocentric beams, with no aberrations). This means that any ray emerging from A comes to A', no matter what the ray angle *u*, and in terms of aberration it is equivalent to zero spherical aberration. Furthermore, the small object, dy, is imaged by the surface Q into dy', both the object and the image being perpendicular to the optical axis and starting in the stigmatic points A and A'. If linear magnification V = dy'/dy is independent of the ray angle, *u*, and remains constant it means that the following relation is valid:

$$V = \frac{n \sin u}{n' \sin u'} = \frac{S'}{S} = \text{const.}$$
(2.24)



FIGURE 2.15 The sine condition.

This relation is known as the sine condition and violation of it is usually called offence against the sine condition (OSC). Obeying the sine condition actually means that the small object dy is imaged with no aberration and the off-axis side of dy' is not blurred (it is a point and not a spot, i.e., there is no coma aberration here).

The pair of conjugate points where the spherical aberration is zero and the sine condition is kept valid are known as aplanatic points of the surface Q. If the object (and the front aplanatic point) is located at infinity the sine condition is transformed into following relation:

$$f' = \frac{h}{\sin u'} \tag{2.25}$$

for any height h at which the ray strikes the surface Q. Then

$$OSC = \delta f' = \frac{h}{\sin u'} - f'.$$

Needless to say, aplanatic points of the surface are of great significance, since in the vicinity of these points imaging occurs with no aberration, even for a beam of a wide solid angle.

A spherical surface has at least three pairs of aplanatic points, two of them being trivial, like the point, C, where the surface crosses the optical axis or the center, O, of the sphere curvature (see Fig. 2.16a). The third aplanatic point pair, A, A' (shown in Fig. 2.16b), is defined by the relations

$$S = CA = \frac{n+n'}{n}r; \quad S' = CA' = \frac{n+n'}{n'}r$$
 (2.26)

where r is the radius of curvature of the surface. Magnification at these points obeys the expression



 $V = \left(\frac{n}{n'}\right)^2 \tag{2.27}$

FIGURE 2.16 Aplanatic points of a spherical refraction surface: (a) in the center of curvature; (b) off-center points.



FIGURE 2.17 Aplanatic points of (a) positive and (b) negative lenses.

while magnification of the center point, O, is

$$V_{\rm O} = \left(\frac{n}{n'}\right). \tag{2.28}$$

Combination of two spherical surfaces with two kinds of aplanatic points enables one to create lenses where imaging is performed with no aberration (theoretically). Two such examples, one of a positive lens and another of a negative one, are depicted in Fig. 2.17 (see also details in Problem 2.11).

Problems

2.10. *A ball lens.* Find the OSC plot for a sapphire ball lens of 3 mm diameter working with an object at infinity and calculate the maximum diameter of the beam which can be concentrated behind the lens. The refractive index of sapphire is 1.77.

2.11. How does one design the aplanatic objective of a microscope if the required magnification is V = -6 at least and it is known that the gap of 0.7 mm between the object plane and the first lens surface is filled with immersion oil of n = 1.8? [Note: The objective should be constructed from two lenses. The first, which is a hemispherical lens, is 1 mm distant from the second lens of 3 mm thickness. The first component is made of SF-57 glass and the glass for the second lens can be chosen using the data of Appendix 2.]

2.1.6. Addition of Aberrations

In practical situations when an imaging system comprises several elements (sometimes consisting of ten or more components), estimating the contribution of each element to the total aberration balance is quite useful. There are several rules allowing one to add aberrations of separate elements and to calculate their impact at different locations along the optical axis. One should keep in mind, however, that the main goal is to reveal how aberrations of each element affect the final image. Addressing the procedure of addition of aberrations we suppose that the *i*-th element (or a group of elements) performs imaging from its object space to the image space with some linear magnification, V_i , and the image built by the *i*-th element serves as a virtual object for the next (*i* + 1)-th element. The following rules should be followed:

• the lateral aberrations $\delta s'_{i-1}$ while being transferred from the object space to the image space of the *i*-th component are multiplied by V_i^2 , so that the total lateral aberration at the image space of the *i*-th element becomes

$$\delta s'_{\text{tot}} = \delta s'_{i-1} \times V_i^2 + \delta s'_i; \qquad (2.29)$$

• the transverse aberrations while being transferred from the object space to the image space of the *i*-th component are multiplied by V_i , so that the total transverse aberration at the image space of this component becomes

$$\delta s'_{t, \text{tot}} = \delta s'_{t, i-1} \times V_i + \delta s'_{t, i}; \qquad (2.30)$$

- if the *i*-th element transfers images to infinity it should be treated as if radiation propagates in the opposite direction and the aberrations computed in such a manner should be added to the aberrations of the image space of the (*i* − 1)-th element (δ^{*k*−'}_{S_i} with its sign, and δ^{*k*−'}_{S_i} with the opposite sign);
- if a parallel beam is created between two subsequent elements they should be considered as a group with a single magnification and aberrations are treated as if they are transferred from the object space of the first element of this group to the image space of the second element of the group;
- addition of aberrations should be done separately for aberrations along the chief ray and for aberrations along the marginal ray.

These simple rules assist in the analysis and synthesis of imaging systems as far as aberrations are concerned.

Problems

2.12. In a two-lens imaging system (Fig. 2.18) initially aligned to get a sharp image of an object of 0.25 mm on a CCD of size 5 mm \times 5 mm, a scale reticle R of 2 mm thickness is introduced in the plane P where the intermediate image is formed at magnification $V_1 = -5$. Both lenses are of 8 mm diameter and 15 mm focal length and they are properly corrected (the lens aberrations can be neglected). Find the impact of the reticle on aberrations in the CCD plane and the way to make a correction.


FIGURE 2.18 Problem 2.12 – Two-lens imaging system with reticle.



FIGURE 2.19 Problem 2.13 - (a) Residual lateral aberration of a lens and layouts with bending by (b) a mirror and (c) a penta-prism.

2.13. A lens of 40 mm in size and f # = 1.2 performs imaging of a distant object to the detector plane P and has the residual aberration shown in the plot of Fig. 2.19a. The optical axis should be turned through 90° and two possible configurations are compared: one with a plane mirror and the other with a penta-prism (see Figs. 2.19b and 2.19c, respectively). What is the advantage of the second layout and what is the optimal size of the prism?

2.14. A two-lens condenser. An illumination system (Fig. 2.20) aiming to concentrate radiation from a halogen lamp with 3 mm filament into an optical fiber bundle of 6 mm in size consists of two lenses: L_1 of 30 mm diameter and 60 mm focal length and L_2 of the same diameter and 120 mm focal length, both made



FIGURE 2.20 Problem 2.14 – Configuration of a two-lens condenser.

of BK-7 glass. Find the optimal shape of the condenser lenses and estimate the spherical aberration at the bundle entrance.

2.2. Diffraction Effects and Resolution

2.2.1. General Considerations

Diffraction effects result from the wave nature of radiation participating in imaging. In general diffraction is caused by the secondary waves generated in the substance of an obstacle on which electromagnetic waves impinge while traveling in space. An obstacle can be a body of any shape, either transparent or opaque. Interference of the secondary waves changes the spatial distribution of the propagated radiation in such a way that light energy appears not only in the direction of the initial propagation but also to the side of it. Because of this, for example, an ideal lens with no aberration is not capable of concentrating light in a single point of the image plane and some energy is always revealed in a small but finite vicinity of the image. Thus, diffraction is a basic limitation in imaging optics which cannot be avoided. Other effects, like aberrations considered in the previous section, which also "spoil" the image quality appear together with diffraction remains a single factor affecting the system performance. In such a case the optical system is termed diffraction limited.

Diffraction occurs at any stop through which light passes. It could be a real aperture, or the mounting of a lens, prism, or mirror, or just the boundaries of an optical element of the system. We shall consider a simple case of propagation of monochromatic light of wavelength λ through a circular non-transparent stop of radius *a* followed by a lens (see Fig. 2.21). It can be shown that the intensity



FIGURE 2.21 Diffraction on (a) a circular stop and (b) the intensity distribution in the diffraction spot.

distribution of light in the spot created in the image plane P due to diffraction is governed by the following function (Airy's function):

$$I(r) = I_0 \left[\frac{2J(x)}{x} \right]^2; \quad \text{where } x = \frac{2\pi}{\lambda} n' r' \sin u'_{\text{max}}, \qquad (2.31)$$

n' is the refractive index in the image space, r' is the radial coordinate in the plane P, u'_{max} is the maximum angle of the direction from the stop boundary to the center of the spot, and $J_1(x)$ is the Bessel function of the first order.

Expression (2.31) is an oscillating function with a strong central maximum followed by dark and light rings of decreasing intensity. It is commonly accepted that most of the energy of the spot is concentrated in the central maximum limited by the first dark ring which corresponds to the value $x_{\min}^{(1)} = 3.8317$ in Eq. (2.31). Hence, the relevant size of the spot in the plane P obeys the relation

$$\delta_{\rm dif} = \frac{1.22\lambda}{n'\sin u'_{\rm max}}.$$
(2.32)

In the case when P is the focal plane of a lens of diameter D = 2a, Eq. (2.32) is transformed into the well-known expression

$$\delta_{\rm dif} = \frac{2.44\lambda}{D} f' \quad (n'=1).$$

The diffraction spot has a direct impact on limiting resolution which is one of the basic features of any imaging system. Consider two very close images in the plane P, each one generating a diffraction spot. If the distance between the two images is large enough the spots are well separated and an observer looking on the image plane P is capable of perceiving them easily. The smaller the distance, the closer the spots, and at some stage they become overlapped. The question is, what is the minimum distance at which two partially overlapping spots are still recognized as two separate objects? Such a minimal distance is called the limiting resolution and it is defined, according to the Rayleigh criteria, as the situation when the minimum of one spot coincides with the maximum of the second. Figure 2.22 demonstrates the situation when two images, one centered at point A' and the other centered at B', are still resolvable. The dotted line in Fig. 2.22b shows the distribution of energy after summation of both spots. The "valley" between the two maxima is about 70% of the maximum intensity (i.e., about 30% reduction of energy).

What is usually important in practical applications is the distance in the object plane between two points A and B corresponding to limiting resolution in the image plane. Referring to Fig. 2.22a, suppose an entrance pupil of size D_p is located at a distance *p* from the object plane. Taking into account that the product $n \times \sin u \times r$ is the system invariant (it remains constant while transferring through



FIGURE 2.22 (a) System resolution and (b) two spots according to the Rayleigh criteria.

each refraction surface) and using Eq. (2.32), one can transform the distance δ_{dif} into the corresponding distance AB and resolvable angle β in the object plane:

$$AB = \frac{1.22\lambda}{nD_{\rm p}}p; \quad \tan\beta = \frac{AB}{p} = \frac{1.22\lambda}{nD_{\rm p}}$$
(2.33)

or, using the expression in angular seconds, $\beta = 120''/D_p$ for $\lambda = 0.5 \,\mu\text{m}$ and n = 1.

If aberrations of the system are significant then the diffraction spot should be considered together with the aberration spot. The common practice is to use the square root rule for getting the total spot as follows:

$$\delta_{\rm sum} = \sqrt{\delta_{\rm dif}^2 + \delta s_{\rm ab}^2}.$$
 (2.34)

Apparently the resolution limit is affected by Eq. (2.34). A simple way to define resolution with the spot enlarged by aberrations is demonstrated in Fig. 2.23. To find the resolution in the object plane in this case one should divide the value δ_{sum} calculated from Eq. (2.34) by the magnification of imaging.



FIGURE 2.23 Resolution limit caused by aberrations and diffraction.

2.2.2. Diffraction Theory of Imaging in a Microscope

Considering microscopic imaging in terms of diffraction allows one to understand the basic limitations existing in this kind of instrument and to determine relations governing the maximum achievable resolution.

In Abbe's theory of the microscope the object is referred to as a transparent diffraction grating (see Chapter 5) of a spatial period d. This approach is based on the assumption that a real object described by an arbitrary intensity distribution function which can be expanded in a Fourier series of separate harmonics is considered as a collection of sine periodic spatial waves, each one acting as a diffraction grating. Being illuminated by a parallel beam, the grating generates several fans of beams corresponding to different diffraction orders. The zero-order beam is concentrated by the microscope objective at the back focal point whereas the other diffraction orders are collected at other points of the same back focal plane (F_{i+1} ; F_i ;...). The aperture stop located in the back focal plane comprises all focused centers of diffracted beams (see Fig. 2.24). Light of each order proceeds further as a divergent fan to the plane of the field stop positioned in the focal plane of the eyepiece. Here the fans overlap and interfere. The resulting fringe pattern with a constant spacing, d', constitutes an image of the initial grating of the object plane and both are related through the system magnification: d'/d = V.

To create the fringe pattern at least two divergent beams and therefore two diffraction orders must be present simultaneously in the aperture stop. The location of the diffraction maxima, F_i , in the focal plane of the objective is dictated by the diffraction grating equation (Eq. (5.18); see Chapter 5): $\sin u_{\text{max}}^{(i)} = i\lambda/d$ for $i = 0, \pm 1, \pm 2, \ldots$ The aperture stop size, D_{as} , is related to the numerical aperture of the system in the object space: $n \sin u_{\text{max}} = D_{\text{as}}/(2f')$. These last two expressions allow one to find the minimum diffraction spacing, d, which can be imaged by the microscope.

Two possible methods of illumination should be considered separately: direct illumination when the zero-order diffraction is focused in a point on the optical axis



FIGURE 2.24 Diffraction in microscope imaging.



FIGURE 2.25 Location of diffraction maxima in an aperture stop: (a) on-axis illumination; (b) oblique illumination.

and oblique illumination when the focus of the zero-order diffraction is located in an off-axis point. Fig. 2.25 illustrates both situations. The limiting condition for direct illumination requires that the zero-order maximum as well as the 1st and the $(-1)^{st}$ order maxima are inside the aperture stop whereas the corresponding limit for oblique illumination can be realized if the zero-order and only one of the first-order diffraction maxima are inside the circle of diameter D_{as} . As can be seen, the limiting resolution is related to the numerical aperture (NA) of the system as

$$d = \frac{\lambda}{n \sin u_{\text{max}}} = \frac{\lambda}{\text{NA}}; \quad d = \frac{\lambda}{2n \sin u_{\text{max}}} = \frac{\lambda}{2\text{NA}}$$
 (2.35)

for direct (on-axis) and oblique illumination, respectively. In the real practice of microscopy illumination is supplied by a wide-angle condenser coming at both direct and oblique directions. It can be shown that in such a case the limiting resolution of the microscope is determined as

$$d = \frac{\lambda}{\text{NA} + \text{NA}_{\text{C}}}$$
(2.36)

where NA_C is the numerical aperture of the condenser.

Problems

2.15. Find the minimum required active diameter of the well-corrected imaging optics for visible wavelengths operating at a working distance of 30 mm and providing a resolution of $0.5 \,\mu$ m.

2.16. A microscope for the visible range is supplied with three objectives: 10×0.25 NA, 40×0.65 NA, and 100×1.2 NA, and a condenser of 0.96 NA. Find the maximum resolution in all three possible configurations.

2.17. A microscope objective of magnification $\times 10$ has a focal length of 16 mm and is operated with an aperture stop of 5 mm diameter. At which angle of oblique illumination should one expect the resolution to be twice that of normal (on-axis) illumination? Which resolution (in the visible) will be available in this case and how will the resolution be changed if the illumination angle is held at 5°?

2.3. Image Evaluation

Evaluation of images is carried out (i) at the design stage when it is checked whether the configuration designed is capable of delivering the system performance requirements; and (ii) at the end of manufacturing when a real system with all the tolerances of component fabrication and assembling is aligned and prepared for final testing. Image evaluation at the design stage is performed theoretically, by analyzing aberrations of the system and also by calculating some integral parameters enabling one to estimate the expected image quality. Image evaluation at the manufacturing stage is done with special hardware allowing one to measure resolution, contrast, and other parameters related to the system performance, usually determined in a procedure specific for each tested architecture.

Theoretical evaluation of image quality is usually based on ray tracing of a great number of rays, originating in on-axis and off-axis points of the object and, if necessary, related to several representative wavelengths (mostly, the lines C, D, and F) of the illuminating radiation. Obviously computing is carried out with special software allowing one to calculate and display the location of the rays in the image plane (a spot diagram), energy distribution in a spot, frequency response of the system (modulation transfer function, see below), position of the best focus, and other useful parameters. Diffraction effects are also taken into account while computing the relevant parameters and convolution between geometrical optics results (ray tracing) and the diffraction pattern at each and every image point is accurately calculated.

Examples of spot diagrams for on-axis and off-axis points are depicted in Fig. 2.26. Each diagram is calculated by tracing the rays striking the entrance pupil as a uniformly distributed fan and indicating the points of intersection of the rays with the image plane. Apparently for a perfect lens the spot diagram is transformed in a single point located in the paraxial image of a corresponding object point. The on-axis spot is usually symmetrical whereas the off-axis spot might be strongly asymmetric (as shown in Fig. 2.26b) which is an indication of strong field aberrations.

The size of the spot, δ , can be used as a simple evaluation parameter. In some cases this value can also be estimated analytically, using expressions for



FIGURE 2.26 Spot diagram for (a) an on-axis point and (b) a field (off-axis) point.

the third-order aberrations (e.g., Eqs. (2.14) and (2.15)) or for the diffraction spot (Eq. (2.32)).

Once the spot diagram is found it is possible to count the number of rays intersecting the image plane inside a circle of a chosen size. With the assumption that each ray bears the same amount of energy such a procedure (performed several times, for circles of different diameter, d) gives another parameter called the "encircled energy distribution" (see Fig. 2.27). The circle diameters can be taken in absolute units or in relative units, in terms of the unit $z = d/\delta_{dif}$, where δ_{dif} is the spot of a diffraction-limited system, as per Eq. (2.32).

The distributions shown in Fig. 2.27c illustrate the action of a perfect lens (curve 1) compared to a real lens with aberrations (curve 2). As can be seen, in the first case there are some oscillations on the graph which are evidently related to the interference rings of Airy's function. Curve 1 can be found analytically. Denoting the relative energy inside the circle $d(E(d)/E_{tot})$ as L(d), one finds the following expression (for details, see Born and Wolf, 1968):

$$L(d) = 1 - J_0^2(z) - J_1^2(z); \quad z = \frac{\pi D_p d}{2\lambda p}$$
(2.37)



FIGURE 2.27 Encircled energy distribution: (a) schematic of rays inside different circles; (b) relative energy distribution vs. circle size in absolute units; (c) relative energy distribution vs. circle size in units of z.

where D_p is the exit pupil size located at a distance *p* from the image plane. The first minimum occurs at z = 3.8317 and the corresponding encircled energy is about 84%. Aberrations influence significantly the energy distribution and therefore this distribution can be used as a tool for image quality evaluation.

The most common way to evaluate images is based on the modulation transfer function (MTF). To explain this approach we consider the basic relation between an object T(x, y) and its image I(x', y') created by an optical system. The system is characterized by the point spread function (PSF) S(x, x', y, y') which is actually the pattern created in the image space resulting from a single point object. Then the image of an arbitrary object T can be represented as follows:

$$I(x', y') = \iint S(x - x', y - y')T(x, y) \, dxdy$$
(2.38)

which is the convolution between the object and the PSF. By performing the Fourier transform of Eq. (2.38) we get

$$R(k_x, k_y) = Q(k_x, k_y) \times H(k_x, k_y)$$
(2.39)

where the term on the left-hand side, called the frequency response in the image space and given by

$$R(k_x, k_y) = \frac{1}{2\pi} \iint I(x', y') \exp[i(k_x x' + k_y y')] dx' dy', \qquad (2.40)$$

is related through the system optical transfer function (OTF),

$$Q(k_x, k_y) = \frac{1}{2\pi} \iint S(x, y) \exp[i(k_x x + k_y y)] \,\mathrm{d}x\mathrm{d}y \tag{2.41}$$

to the harmonics of the object, $H(k_x, k_y)$. In the above expressions $k = 2\pi/v$, where v is the spatial frequency in cycles/mm.

Since the OTF is generally a function in a complex space characterized by its amplitude and phase, it is valuable to consider its modulus (a real function) called the MTF. MTF(v) = |Q(v)|. The MTF does not depend on the object, but only on the system properties and this is the reason why it is widely used for image quality evaluation. The MTF also can be interpreted in terms of modulation, which is a feature related to the intensity of light and can be easily measured in practice. Let the light intensity in the object vary from I_{max} to I_{min} . Then the contrast revealed in the object plane can be characterized by the modulation M_0 :

$$M_{\rm o} = \frac{I_{\rm max} - I_{\rm min}}{I_{\rm max} + I_{\rm min}} \tag{2.42}$$

and the contrast in the image plane is described by the corresponding modulation M_i . The ratio between the two modulations is governed by the MTF:

$$M_{\rm i}(v)/M_{\rm o}(v) = \rm MTF(v) \tag{2.43}$$

if all three values are determined for the same spatial frequency v.

Another attractive feature of the MTF is that the MTF of an imaging system is just the product of the MTFs of the separate components constituting the system. This results from linearity and other features of the Fourier transform. Thus, adding or replacing an element can be easily analyzed with regard to the new image quality.

Computation of MTF(v) is a cumbersome and time-consuming procedure which in most cases is performed by special software. However, there are a few cases when it can be expressed explicitly, in analytical form. For example, for a diffraction-limited system, with no aberration, the PSF is Airy's function described by Eq. (2.31). Its OTF and MTF can be found analytically as follows (e.g., see Smith, 1984):

$$MTF(v) = \frac{2}{\pi} (F - \sin F \cos F) (\cos \beta)^n; \quad F = \arccos\left(\frac{\lambda v}{2NA}\right)$$
(2.44)

where β is half of the full-field angle, NA is the numerical aperture in the image space (NA = $n' \sin u'$), and the power n = 1 or n = 3 for radial or tangential directions, respectively. Obviously MTF = 0 at F = 0, meaning that the spatial frequency

$$v_{\rm c} = \frac{2\rm NA}{\lambda} \tag{2.45}$$

is the maximum frequency transferred by the system from the object to the image space (it is called the cut-off frequency).

Figure 2.28 illustrates theoretically calculated MTFs. The diffraction-limited system (curve 1) features the highest MTF at any spatial frequency. It can also



FIGURE 2.28 Modulation transfer functions for a diffraction-limited system (1) and for systems with aberrations (2 and 3).



FIGURE 2.29 (a,c) Input square waves and (b,d) the corresponding output patterns.

be seen that the influence of aberrations is more significant at higher frequencies (difference between curve 1 and curves 2 and 3). Curve 3 also shows that in some cases the MTF might have negative values, which means a 180° phase inversion (black zones become white and vice versa).

According to the explanation above, the MTF (and OTF), strictly speaking, refers to the harmonics, or sine waves, in the object space. In reality, however, the same approach is also exploited for "square wave" objects, like a bar code. As is demonstrated in Fig. 2.29, the contrast and the modulation M_i in the image plane decrease when the spacing (period) of the object square wave, T(x), decreases.

Using a target bar code with several groups of well-defined spatial frequencies as the system object and measuring the modulation M_i of the corresponding images at the system output allows one to find the MTF (see Eq. (2.43)). This method is commonly exploited in image quality evaluation at the final testing stage. The limiting resolution of the system is defined as the spatial frequency of the group still visible at the image plane with a minimum contrast of 3–5% (which is considered as the limit of the perception capability of a human eye).

Problems

2.18. What could be concluded about imaging optics if an analysis of the encircled energy revealed 45% of the total energy of the spot corresponding to an on-axis image point is inside the circle diameter which is half the size of the whole spot?

2.19. Imaging optics operated with monochromatic illumination of 0.6 μ m and having NA = 0.25 in the object space is ended by a CCD area sensor. What should be the minimum pitch of the CCD in order to acquire all spatial frequencies transferred by the optics?

2.20. MTF measurements are carried out with a square-wave target of variable frequencies made of chrome on glass in a bar code pattern. Data are collected for low spatial frequency ($v_1 = 10$ cycles/mm) and for high spatial frequency ($v_2 = 200$ cycles/mm) and the respective measured modulations are 70% and 20%. Assuming that the reflectance of chrome is 70% and the reflectance of glass is 4% and also keeping in mind that the contrast of images is slightly degraded by the background light scattered inside the measurement set-up, find the true MTF value at higher spatial frequency.

2.21. A diffraction-limited optical system operated in the visible range and having NA = 0.15 creates an image on a CCD sensor followed by a video monitor. The MTF of CCD + monitor is 60%. Could we expect to see on the screen the tiny details of an object corresponding to the spatial frequency of 575 cycles/mm?

2.4. Two Special Cases

2.4.1. Telecentric Imaging System

This kind of architecture is usually exploited in measurement systems where errors caused by the third dimensions (along the optical axis) of an object have to be minimized. To explain this error (sometimes called the parallax error, or perspective error) we refer to Fig. 2.30a where simple imaging with a single lens is depicted. Two objects, O_1A and O_2B , having the same height and located at different distances from the lens, after imaging are transformed into images O'_1A' and O'_2B' of different heights. The error $\Delta y'$ might cause problems if the defocusing $\Delta x'$ is small (not revealed by the system observer). The telecentric imaging system shown in Fig. 2.30b is free of this error. The system is configured as an afocal lens



FIGURE 2.30 Imaging with (a) parallax error and (b) the telecentric configuration.

pair where the back focal plane of the first lens coincides with the front focal plane of the second. What is also important is that the aperture stop ab is located in this plane P. As a result, the entrance pupil and the exit pupil are both located at infinity (one on the object space side and the other on the image space side) and the chief rays originating in points A and B are parallel to the optical axis in both spaces. Hence the images O'_1A' and O'_2B' are of the same size and the parallax error does not occur.

If the system is operated with a video area sensor (like a CCD) it should be positioned in such a way that both images are sharp enough. Aberration of defocusing (like the other aberrations) strongly depends on the active lens size, but a reduction in the lens diameter is accompanied by the increasing impact of diffraction, as discussed in Section 2.2, and also a decrease in the image illumination. Therefore, a compromise should be found. In any case, symmetrical configurations are preferred where the shapes of the lenses are equally positioned with regard to the plane P (or one of them is scaled in a symmetrical manner, if magnification/minification is required). Estimation of aberrations can be carried out by the method described in Section 2.1.6 and Problem 2.14.

2.4.2. Telephoto Lens

There are numerous situations where the effective focal length of the objective has to be long while the actual size of the lens should be kept as small as possible. A possible architecture in such a case is a two-lens configuration, one of positive and the other of negative optical power (see Fig. 2.31). Usually what is known is the equivalent focal length, f'_{e} , and the desired length of the configuration, *l*. The optical power of each component and their locations with regard to the image plane should be found.



FIGURE 2.31 Configuration of a telephoto lens.

Considering the system in terms of first-order optics (paraxial approximation) we have for this two-lens system (see Problem 1.7)

$$\Phi = 1/f'_e = \Phi_1 + \Phi_2 - \Phi_1 \Phi_2 d \tag{2.46}$$

and taking into account that $d + S'_F = l$ we also get

$$\Phi_1 d = 1 - \Phi(l - d). \tag{2.47}$$

Equations (2.46) and (2.47) for three unknowns, Φ_1 , Φ_2 , and d, allow one to introduce an additional condition to optimize the configuration with regard to aberration. This could be either the requirements for a minimum optical power of the second element (which in general might result in lower residual aberrations) or the requirements for a configuration with minimal (better zero) curvature of the image surface. In the first case the best results, as can be shown, are obtained with d = 0.5l and the corresponding focal lengths of the elements are

$$f_1' = \frac{lf_e'}{2f' - l_e}; \quad f_2' = -\frac{l^2}{4(f' - l)}.$$
(2.48)

In the second case a zero Petzval's sum (see Section 2.1.4) is required which is achieved with $\Phi_2 = -\Phi_1$. By introducing this condition in Eqs. (2.46) and (2.47) we have

$$l = 0.75f'_{\rm e}; \quad f'_1 = -f'_2 = 0.5f'_{\rm e}; \quad S'_{\rm F} = 0.5f'_{\rm e}.$$
 (2.49)

The latter approach is widely used in the design of telephoto lenses intended for imaging in large angular fields of view.

Problems

2.22. How does one design a telecentric imaging system which is operated at magnification V = -3 in an angular field of view of $\pm 5^{\circ}$ and provides a resolution of 2 μ m in the visible spectral interval?

[Note: Assume the system is free of aberration.]

2.23. A telephoto lens forms images with negligible curvature at a distance of 60 mm from the first (front) element. What are the focal lengths and the distance between the lenses?

2.5. Solutions to Problems

2.1. Since the lens is working in the paraxial range (f # = 10) we can find the distance to the plane P where an ideal image is formed by the lens with nominal



FIGURE 2.32 Problem 2.1 - (a) Defocusing in the plane P and (b) the meridional aberration plot.

focal length:

$$\frac{1-V}{S'} = \frac{1}{f'}; \quad S' = 100 \times (1+2) = 300 \text{ mm}.$$

If the lens is manufactured with 5% tolerance the focal length might be as long as 315 mm, and in this case blurring due to defocusing occurs in the plane P, as demonstrated in Fig. 2.32a. The maximum lateral aberration of defocusing, $\delta s'_l = 15$ mm, gives the corresponding transverse aberration calculated as in Eq. (2.1): $\delta s'_t = \rho_{\text{max}} = \delta s'_l \times \tan u_{\text{max}} = 15 \times 5/315 = 0.238$ mm. Obviously this aberration is a linear function of the height, y, of the ray at the entrance pupil. In the plot shown in Fig. 2.32b the relative vertical coordinate is exploited, y/h_{max} , which varies in the range from 1.0 to (-1.0).

2.2. We start by calculating the thickness of the glass block equivalent to the penta-prism (see Section 1.4). We have $t_e = 3.41a = 3.41 \times 30 = 102.43$ mm and the prism makes the optical path of a chief ray longer by $t_e(1 - n)/n = 102.4(1 - 1.5163)/1.5163 = 34.87$ mm. Figure 2.33 demonstrates the divergent beam traveling through the slab of thickness t_e and explains the appearance of lateral aberration $\delta s'$ as a function of the incidence angle u. Using the rigorous formula from Problem 1.6, one can find the aberration $\delta s'$ as a difference between lateral segments calculated for the paraxial range and for any final angle u:

$$\delta s' = t_{e} \left[(1 - 1/n) - \left(1 - \sqrt{\frac{1 - \sin^2 u}{n^2 - \sin^2 u}} \right) \right] = \frac{t_{e}}{n} \left(1 - n\sqrt{\frac{1 - \sin^2 u}{n^2 - \sin^2 u}} \right)$$

For the maximum angle defined as $\tan u_{\text{max}} = 20/125$ we get from the above equation $\delta s' = 0.48$ mm and for half of the maximum height we obtain $\tan u = 10/125$; $\delta s' = 0.11$ mm. The final plot of aberration introduced by the prism is presented in Fig. 2.33b.



FIGURE 2.33 Problem 2.2 - (a) Geometry of rays traveling through an unfolded prism and (b) the aberration plot.



FIGURE 2.34 Problem 2.3 – Chromatism of a penta-prism: (a) the ray diagram; (b) the aberration plot.

2.3. Proceeding with the penta-prism considered in Problem 2.2, we refer to Fig. 2.34a and calculate the displacement segment AA' = L separately for three main wavelengths, C, D, and F. Chromatic aberration is determined as $L_c - L_d$ and $L_F - L_D$.

Keeping in mind that for BK-7 glass (see Appendix 2) $n_D = 1.5168$, $n_C = 1.51432$, and $n_F = 1.52238$, and starting with the simplified expression for *L* valid in the paraxial range we have

$$\delta s_{\rm Ch}' = L_{\rm C} - L_{\rm D} = t_{\rm e} \left[\left(1 - \frac{1}{n_{\rm C}} \right) - \left(1 - \frac{1}{n_{\rm D}} \right) \right] = 102.43 \left(\frac{1}{1.5168} - \frac{1}{1.51432} \right)$$
$$= -0.111 \text{ mm.}$$

Before proceeding further, we compare the result obtained above with the calculation by the rigorous formula for the maximum angle of incidence, $i_{\text{max}} = \arctan(10/125) = 9.09^{\circ}$:

$$\delta_{\rm Ch}' = t_{\rm e} \left[\left(1 - \sqrt{\frac{1 - \sin^2 i_{\rm max}}{n_{\rm C}^2 - \sin^2 i_{\rm max}}} \right) - \left(1 - \sqrt{\frac{1 - \sin^2 i_{\rm max}}{n_{\rm D}^2 - \sin^2 i_{\rm max}}} \right) \right] = -0.112 \text{ mm}.$$

As we see, the difference between exact solution and the paraxial approximation is very small in our case, so that we proceed with the simplified formula and calculate

$$\delta s'_{\rm Ch}^{(2)} = L_{\rm F} - L_{\rm D} = t_{\rm e} \left[\left(1 - \frac{1}{n_{\rm F}} \right) - \left(1 - \frac{1}{n_{\rm D}} \right) \right]$$
$$= 102.43 \left(\frac{1}{1.5168} - \frac{1}{1.52238} \right) = 0.248 \text{ mm}.$$

The plot of the chromatic aberration of the prism is shown in Fig. 2.34b.

2.4. Using the glass data from Appendix 2, we get for the components of the doublet lens (achromat) the following (see Eq. (2.11)):

$$\Phi_1 = \Phi \frac{v_{\text{D1}}}{v_{\text{D1}} - v_{\text{D2}}} = 0.01333 \frac{64.12}{64.12 - 33.686} = 0.028084$$
$$\Phi_2 = -\Phi \frac{v_{\text{D2}}}{v_{\text{D1}} - v_{\text{D2}}} = 0.01333 \frac{33.686}{64.12 - 33.686} = -0.014751.$$

Assuming the first surface is plane $(r_1 = \infty)$, we can find the second radius from Eq. (1.11) for a thin lens: $r_2 = -(0.5168/0.028084) - 18.40$ mm and this is also the first radius (r_3) of the second element. Hence the last radius can be calculated as follows:

$$\frac{1}{r_4} = \frac{1}{r_3} - \frac{\Phi_2}{n_{D2} - 1} = -\frac{1}{18.4} + \frac{0.014751}{0.62588} = -0.0307794; \quad r_4 = -32.489 \text{ mm.}$$

To find the residual aberration we calculate the focal length of the doublet in all three main wavelengths using Eq. (2.8). We have for the positive component

$$\Phi_{1C} = \frac{0.51432}{18.40} = 0.027952; \ \Phi_{1D} = 0.028084; \ \Phi_{1F} = \frac{0.52238}{18.40} = 0.02839$$

and for the negative component:

$$\Phi_{2C} = -0.62074 \left(\frac{1}{18.40} - \frac{1}{32.49} \right) = -0.14630; \quad \Phi_{2D} = -0.014751;$$

$$\Phi_{2F} = -0.63932 \times 0.023568 = -0.015067.$$



FIGURE 2.35 Problem 2.4 – (a) The doublet lens and (b) its secondary spectrum.

Then

 $\Phi_{\rm C} = \Phi_{1\rm C} + \Phi_{2\rm C} = 0.027952 - 0.014630 = 0.013322; \quad f_{\rm C}' = 75.06 \text{ mm}$ $\Phi_{\rm F} = \Phi_{1\rm F} + \Phi_{2\rm F} = 0.02839 - 0.015067 = 0.013323; \quad f_{\rm F}' = 75.06 \text{ mm}$

and the residual aberration (secondary spectrum) is $\delta s'_{Ch} = f'_C - f'_D = 0.06 \text{ mm}$ (see the plot depicted in Fig. 2.35b).

2.5. From the problem data it follows that the focal length of the lens is $f' = f\# \times D = 60$ mm.

(a) Starting with the lens made of BK-7 glass and data from Appendix 2 we have $\xi = 1/n_D = 1/1.5168 = 0.6593$. By substituting this value in Eq. (2.16) we get the radii of the lens of the optimal shape:

$$r_1 = 2(1 - 0.6593) \frac{2.319}{2.6593} 60 = 35.65 \text{ mm};$$

$$r_2 = \frac{2 \times 0.3407 \times 2.319}{2 - 0.6593 - 4 \times 0.6593^2} = -238.2 \text{ mm}$$

and the lateral spherical aberration at the maximum height h = D/2 = 15 mm, as per Eq. (2.15), is

$$\delta s'_{\text{Sph}} = -\frac{1}{8} \times \frac{0.6593}{0.3407^2} \times \frac{3.3405}{2.319} \times \frac{15^2}{60} = 3.84 \text{ mm}.$$

This yields the transverse spherical aberration as $\delta s'_t = \delta s'_{\text{Sph}} \times 15/60 = 0.96 \text{ mm.}$

If the lens is made of SF-11 glass the optimal shape is different. Doing just as above, but with $n_D = 1.78472$, we obtain $r_1 = 2 \times 0.43969 \times 2.1206 \times (60/2.5603) = 43.7$ mm and the corresponding value for the second radii becomes

$$r_2 = 2 \times 0.43969 \times 2.1206 \times \frac{60}{2 - 0.5603 - 4 \times 0.5603^2} = 608.2 \text{ mm.}$$

As we see, the second radius in this case is also positive so that the optimal shape is a meniscus with a very large second radius. The value of the lateral spherical aberration, if the lens stands optimally, is

$$\delta s'_{\text{Sph}} = -\frac{1}{8} \times \frac{0.5603}{0.43969^2} \times \frac{3.4397}{2.1206} \times \frac{15^2}{60} = 2.20 \text{ mm}$$

which gives the transverse aberration as $\delta s'_t = 0.55$ mm.

(b) If the object is at infinity, but the lens does not stand optimally (e.g., the second radius, r_2 , is directed to the object, meaning that the lens is turned by 180°), the calculations should be based on the more general formula (Eq. (2.14)) which gives for the BK-7 lens

$$\delta s'_{\text{Sph}} = -\frac{1}{2} \times \frac{225}{0.3407^2} \left(\frac{1}{60} - \frac{2.6593 \times 0.3407}{238.2} + 2.319 \times 60 \times \frac{0.3407^2}{238.2^2} \right)$$

= -12.75 mm

and $\delta s'_t = -3.19$ mm. For the SF-11 lens $\delta \dot{s}'_{Sph} = -10.77$ mm and $\delta s'_t = -2.69$ mm. As we see, optimization of the lens position causes an about three times reduction of the spherical aberration for the BK-7 lens and an about five time reduction for the SF-11 lens.

2.6. (a) Let us find first the paraxial parameter of a cylinder lens (a rod). Since $r_1 = -r_2 = 7/2 = 3.5$ mm and d = 7 mm, we obtain from the lens formula (Eq. (1.11))

$$\frac{1}{f'} = (1.5168 - 1)\frac{2}{3.5} - \frac{7 \times (1.5168 - 1)^2}{3.5^2 \times 1.5168} = 0.1947; \quad f' = 5.136 \text{ mm}$$

and using this value in Eq. (1.10) we find the location of the principal planes:

$$a' = f' - S'_{\rm F} = \frac{5.136 \times 7 \times 0.5168}{3.5 \times 1.5168} = 3.5 \,\,{\rm mm}.$$

Therefore, both principal planes are located in the center of the lens and the distance to the object is S = -(3.5 + 2) = -5.5 mm. Then the distance to the paraxial image is $S' = (1/S + 1/f')^{-1} = (1/5.136 - 1/5.5)^{-1} = 77.83$ mm, i.e., the plane P passing through the paraxial image is located at a distance $l'_0 = 74.33$ mm from the lens. It is the plane P where we should calculate the transverse spherical aberrations.

Before proceeding further, we will consider the general case of ray tracing through a full cylinder lens. Referring to Fig. 2.36, we find the segment Δ_1 from two triangles, AA₁B and OA₁B: $\Delta_1 = \rho - \sqrt{\rho^2 - y_1^2} = (y_1/\tan u) - l$, where $\rho = D/2$ is the radius of the lens. By dividing both sides by ρ and denoting

$$a = 1 + l/\rho; \quad b = 1/\tan u; \quad \sin \varphi = z$$
 (A)



FIGURE 2.36 Problem 2.6 – Ray tracing through a cylindrical rod.

we have the following equation with regard to z:

$$1 - z^2 = a^2 - 2abz + b^2 z^2.$$
 (B)

Solving this one can find the angle φ as follows:

$$\sin \varphi = z = \frac{ab - \sqrt{a^2 b^2 - (a^2 - 1)(b^2 + 1)}}{b^2 + 1}.$$
 (C)

This yields further $i_1 = u + \varphi$; $r_1 = \arcsin(\sin i_1/n)$ and from the triangle A₁OA₂, taking into account that $r_1 = r_2$; and $i_1 = i_2$, we have $180^\circ = \varphi + \beta + (180^\circ - 2r_1)$ which gives

$$\beta = 2r_1 - \varphi; \quad \gamma = i_1 - \beta. \tag{D}$$

As $y_2 = \rho \sin \beta$ and $\Delta_2 = \rho - \rho \cos \beta$, we finally obtain

$$l' = \frac{y_2}{\tan \gamma} - \Delta_2 = \rho \left(\frac{\sin \beta}{\tan \gamma} - 1 + \cos \beta \right).$$
(E)

The lateral spherical aberration for each ray angle u is found as the difference between l'_0 and the value l' from Eq. (E). The corresponding transverse spherical aberration is determined as

$$\delta s'_{t} = \tan \gamma \times (l' - l'_{0}). \tag{F}$$

To build the aberration plot we should repeat the procedure as per Eqs. (A)–(F) for several angles *u*. The whole range of *u* can be found as the following: $\sin u_{\text{max}} = \rho/(l + \rho)$ and according to the data of the problem we find $u_{\text{max}} = 39.52^{\circ}$. If the entrance pupil is located at the distance l = 2 mm from the light source (the object)

the maximum coordinate in the pupil plane is $y_{max} = l \times \tan 39.52 = 1.65$ mm and the relative coordinate of a point in the entrance pupil is calculated as $\tilde{y} = y/1.65$.

We start with $u_1 = 35^{\circ}$. This gives $y_{01} = 2 \tan 35^{\circ} = 1.4009$; $\tilde{y}_1 = 0.849$. Then from Eqs. (A)–(E) we find step-by-step: a = 1.5714; b = 1.428; $z = \sin \varphi = 0.490$; $\varphi = 29.33^{\circ}$; $i_1 = u_1 + \varphi = 64.33^{\circ}$; $\sin r_1 = 0.5942$; $r_1 = 36.46^{\circ}$; $\beta = 43.58^{\circ}$; $\gamma = 20.75^{\circ}$; l' = 5.403 mm. This leads to $\delta s'_{\text{Sph}} = 5.403 - 74.33 = -68.93$ mm; $\delta s'_{\text{t}} = -\tan 20.75^{\circ} \times 68.93 = -26.11$ mm.

The properties of the second second

Choosing $u_3 = 10^\circ$ we get $y_{03} = 2 \tan 10^\circ = 0.3527$; $\tilde{y}_1 = 0.214$. Then a = 1.5714; b = 5.671; $z = \sin \varphi = 0.1017$; $\varphi = 5.835^\circ$; $i_1 = u_1 + \varphi = 15.835^\circ$; $\sin r_1 = 0.1798$; $r_1 = 10.36^\circ$; $\beta = 14.893^\circ$; $\gamma = 0.942^\circ$; l' = 54.59 mm. This leads to $\delta s'_{\text{Sph}} = 54.59 - 74.33 = -19.73$ mm; $\delta s'_{\text{t}} = -\tan 0.942^\circ \times 19.73 = -0.324$ mm.

Finally the aberration plot is as shown in Fig. 2.37 (we take into account that the rays incident on the lower half of the entrance pupil are going up at the exit, so that the transverse aberrations here are positive and the diagram is anti-symmetric).

(b) Let the lens be a full sphere (a ball lens) and the object is an on-axis point. Tracing any ray coming to the lens, we consider the plane of incidence which includes the ray and the optical axis. Such a plane is a meridional plane and it is similar to the one shown in Fig. 2.36 for the cylindrical rod. Therefore Eqs. (A)–(F) remain valid and the results will be identical to those obtained in (a) above.

2.7. As was done in Problem 2.6, we find the parameters of paraxial imaging. Referring to Fig. 2.9, we find the focal length and location of the principal planes



FIGURE 2.37 Problem 2.6 – Transverse spherical aberration plot.

of a new lens from Eqs. (1.9)–(1.11), keeping in mind that $r_1 = 3.5$ mm and $r_2 = \infty$:

$$\frac{1}{f'} = \frac{0.5168}{3.5}; \quad f' = 6.772 \text{ mm};$$
$$a' = f' \frac{d(n-1)^2}{r_1 n} = 6.772 \frac{4 \times 0.5168^2}{3.5 \times 1.5168} = 2.637 \text{ mm}$$

To get the paraxial image in the plane P it is necessary to locate the object A at the distance

$$S = -\left(\frac{1}{S'} - \frac{1}{f'}\right)^{-1} = -\left(\frac{1}{74.33 + 2.637} - \frac{1}{6.772}\right)^{-1} = -7.43 \text{ mm}$$

from the front surface of the lens. We choose also the entrance pupil to be 7.43 mm distant from A and assume its size $D_p = 4.4$ mm (to ensure that any ray incident on the entrance pupil will pass the lens and proceed to the plane P). We calculate the aberration plot by performing rigorous ray tracing (see Fig. 2.38). Consideration of the first (bended) surface gives exactly the same expressions (A)–(C) as in Problem 2.6 and the same formulas for the angles i_1 and r_1 :

$$i_1 = u + \varphi; \quad r_1 = \arcsin\left(\frac{\sin i_1}{n}\right).$$

The rest is different. That is,

$$r_{2} = \varphi - r_{1}; \quad y_{2} = \rho \sin \varphi - [t - \rho(1 - \cos \varphi)] \tan(r_{2});$$
$$\gamma = \arcsin(n \sin r_{2}); \quad l' = \frac{y_{2}}{\tan \gamma}.$$
(G)

Once the value of l' is found the transverse aberration is calculated from Eq. (F) of Problem 2.6.



FIGURE 2.38 Problem 2.7 – Transverse spherical aberration plot (plano-convex cylindrical lens).

We apply the above approach choosing first $u_1 = 15^\circ$. We get step-by-step: $y_{01} = 7.43 \times \tan 15^\circ = 1.99$; $\tilde{y}_1 = 1.99/2$. 2 = 0.88. Then a = 3.123; b = 3.732; $z = \sin \varphi = 0.6284$; $\varphi = 38.43^\circ$; $i_1 = u_1 + \varphi = 53.43^\circ$; $\sin r_1 = 0.4143$; $r_1 = 31.97^\circ$; $r_2 = 38.43 - 31.97 = 6.46^\circ$; $\gamma = 9.824^\circ$; $y_2 = 1.833$; l' = 10.58. Therefore, $\delta s'_{\text{sph}} = 10.58 - 74.33 = -63.74 \text{ mm}$; $\delta s'_{\text{t}} = -11.03 \text{ mm}$.

Choosing then $u_2 = 10^\circ$ we have $y_{01} = 7.43 \times \tan 10^\circ = 1.31$; $\tilde{y}_1 = 1.31/2.2 = 0.596$. Then a = 3.123; b = 5.671; $z = \sin \varphi = 0.388$; $\varphi = 22.83^\circ$; $i_1 = u_1 + \varphi = 32.48^\circ$; $\sin r_1 = 0.3574$; $r_1 = 20.94^\circ$; $r_2 = 22.83 - 20.94 = 1.89^\circ$; $\gamma = 2.86^\circ$; $y_2 = 1.235$; l' = 24.70; and $\delta s'_{\text{Sph}} = -49.63$ mm; $\delta s'_t = -2.48$ mm.

For the angle $u_2 = 5^\circ$ we obtain $y_{01} = 7.43 \times \tan 5^\circ = 0.65$; $\tilde{y}_1 = 0.65/2.2 = 0.295$. Then a = 3.123; b = 11.43; $z = \sin \varphi = 0.1873$; $\varphi = 10.795^\circ$; $i_1 = u_1 + \varphi = 15.795^\circ$; $\sin r_1 = 0.180$; $r_1 = 10.338^\circ$; $r_2 = 10.795 - 10.338 = 0.457^\circ$; $\gamma = 0.694^\circ$; $y_2 = 0.626$; l' = 51.68; and finally $\delta s'_{\text{Sph}} = -22.65$ mm; $\delta s'_t = -0.274$ mm.

The resulting aberration plot is presented in Fig. 2.38. Comparing this diagram with that of Fig. 2.37 makes very clear that the full rod has a much greater aberration than the plano-convex lens.

2.8. Assuming the lens performs ideal imaging we find the location of the object and the image from paraxial relations:

$$S = \frac{1-V}{V}f' = -\frac{3}{2}30 = -45$$
 mm; $S' = 90$ mm.

The right-angle prism located 20 mm behind the lens can be unfolded as a parallel glass slab of thickness $t_e = 30$ mm (the dotted line in Fig. 2.39a). So, we can estimate the astigmatism introduced by the prism using Eq. (2.20) for the plate with $d = t_e$. Doing the calculation for several incidence angles $u < 30^\circ$ we find

$$u = 10^{\circ}: u' = \arcsin(\sin 10^{\circ}/1.5168) = 6.574^{\circ};$$

$$\delta'_{Ast} = (1.5168^{2} - 1)\frac{\tan^{3} 6.574}{\tan 10^{\circ}} 30 = 0.338 \text{ mm}$$

$$u = 20^{\circ}: u' = \arcsin(\sin 20^{\circ}/1.5168) 13.03^{\circ};$$

$$\delta'_{Ast} = (1.5168^{2} - 1)\frac{\tan^{3} 13.03^{\circ}}{\tan 20^{\circ}} 30 = 1.32 \text{ mm}$$

$$u = 30^{\circ}: u' = \arcsin(\sin 30^{\circ}/1.5168) = 19.25^{\circ};$$

$$\delta'_{Ast} = (1.5168^{2} - 1)\frac{\tan^{3} 19.25^{\circ}}{\tan 30^{\circ}} 30 = 2.88 \text{ mm}.$$

The astigmatism plot is shown in Fig. 2.39b.



FIGURE 2.39 Problem 2.8 – (a) Layout of the imaging system and (b) astigmatism and, (c) distortion of the right-angle prism.

b)

As to distortion, we have to find first the location of the exit pupil. Since imaging is done by the lens only and the prism is big enough and does not affect the angular field of view, we can assume that the lens itself is the aperture diaphragm and the exit pupil. Then the distance p appearing in Eq. (2.23) is equal to 90 mm. Calculation for several incident angles, u, using Eq. (2.23), yields

for the angle
$$u = 10^{\circ}$$
, $\Delta = \frac{1.5168^2 - 1}{1.5168^3} \times \left(\frac{10}{180}\pi\right)^2 = 0.189\%$
for the angle $u = 20^{\circ}$, $\Delta = 0.76\%$
for the angle $u = 30^{\circ}$, $\Delta = 1.71\%$.

The distortion plot is depicted in Fig. 2.39c.

a)

2.9. We should address here the curvature of field, and, more specifically, check Petzval's condition, as in Eq. (2.21). To find the radii of the lens we use the lens formula in the paraxial range, remembering that the lens is symmetrical $(r_1 = -r_2)$:

$$r_1 = 2f'(n-1) = 2 \times 40 \times 0.5168 = 41.34 \text{ mm} = -r_2.$$

Now Petzval's sum is as follows:

$$\frac{1}{\rho} = -\frac{1}{41.34} \left(\frac{1}{1.5168} - 1 \right) + \frac{1}{41.34} \left(1 - \frac{1}{1.5168} \right) = 0.01648$$

c)



FIGURE 2.40 Problem 2.9 – The flattener and the image plane.

and the curvature of Petzval's surface is $\rho = 60.67$ mm, i.e., the image will suffer significant degradation since off-axis areas have sharp images not in the plane P but rather on the surface of curvature ρ .

The flattener is a negative lens, usually plano-concave, positioned with its plane surface just in the image plane (in the plane P in our case) in a manner demonstrated in Fig. 2.40. Because the flattener practically coincides with the image it does not affect the image magnification. On the other hand, its contribution to Petzval's sum might improve the field curvature. Denoting the first radius of the flattener as r_3 , we rewrite Eq. (2.21) in the following manner:

$$\frac{1}{\rho} = -\frac{2}{r_1} \left(\frac{1}{n} - 1 \right) + \frac{1}{r_3} \left(\frac{1}{n_{\rm Fl}} - 1 \right)$$

where n_{Fl} is the refractive index of the flattener glass. Optimization of r_3 might cause zero curvature ($\rho \rightarrow \infty$) of Petzval's sum (i.e., the image plane becomes flat). This occurs if r_3 is as follows:

$$r_3 = \frac{1/n_{\rm Fl} - 1}{2(1 - 1/n)}r = \frac{1/1.78477 - 1}{2(1 - 1/1.5168)}41.34 = -26.67 \,\,\rm{mm}$$

2.10. To calculate the OSC we use the formulas for rigorous ray tracing derived in Problem 2.6. For the case relevant here when the incident rays are coming from infinity and therefore u = 0 we rewrite the expressions of Problem 2.6 as follows:

$$i_1 = \varphi = \arcsin(h/\rho);$$
 $r_1 = \arcsin(\sin \varphi/n);$ $\beta = 2r_1 - \varphi;$ $\gamma = 2(\varphi - r_1)$
(A)

where *h* is the height of the ray striking the lens and ρ is the ball radius. The OSC for any chosen value *h* is calculated from Eq. (2.25) by substituting the angle *u'* with γ found from Eq. (A) and the focal length, *f'*, as per Eq. (1.11) in a complete form:

$$\frac{1}{f'} = \frac{2(n-1)}{\rho} - \frac{2\rho(n-1)^2}{\rho^2 n} = \frac{2 \times 0.77}{1.5} - \frac{2(0.77)^2}{1.5 \times 1.77} = 0.580;$$

$$f' = 1.724 \text{ mm}.$$



FIGURE 2.41 Problem 2.10 - (a) OSC of a sapphire ball lens and (b) definition of the maximum ray height.

For h = 0.25 mm: $\sin \varphi = 0.25/1.5 = 0.1667$; $\varphi = 9.59^{\circ}$; $r_1 = 5.40^{\circ}$; $\gamma = 8.38^{\circ}$; and

$$OSC = \frac{0.25}{\sin 8.38^{\circ}} - 1.724 = -8.59 \times 10^{-3} \text{ mm.}$$

For h = 0.5 mm: $\sin \varphi = 0.5/1.5 = 0.333$; $\varphi = 19.47^{\circ}$; $r_1 = 10.854^{\circ}$; $\gamma = 17.232^{\circ}$; and

$$OSC = \frac{0.5}{\sin 17.232^{\circ}} - 1.724 = -0.0362 \text{ mm.}$$

For h = 0.75 mm we get in a similar manner OSC = -0.0822 mm and for h = 1.0 mm OSC = -0.147 mm. The plot of OSC is presented in Fig. 2.41a. The horizontal coordinate on the diagram, $\tilde{y} = h/h_{\text{max}}$, is defined as the ratio of the real height to the maximum possible height dictated by the refractive index *n*. To determine h_{max} we refer to the limiting situation shown in Fig. 2.41b. Considering geometry of the ray we obtain

$$\varphi = 2r_1;$$
 $\frac{\sin \varphi}{\sin r_1} = \frac{\sin \varphi}{\sin \varphi/2} = 2\cos \frac{\varphi}{2} = n$

and further $\varphi = 2 \arccos(n/2) = 2 \arccos(1.77/2) = 55.5^\circ$; $h_{\text{max}} = 1.5 \sin 55.5^\circ = 1.236$ mm. Therefore, the maximum diameter of the beam which is concentrated by the lens somewhere behind the ball is 2.472 mm. The rest of the rays striking the ball cross the optical axis inside the lens and cannot be exploited for imaging or any other application related to energy concentration.

2.11. The design of the objective will be based on aplanatic points of spherical surfaces. Choosing the concept depicted in Fig. 2.42, we start with the relations for the first component. Since the immersion oil has practically the same refractive



FIGURE 2.42 Problem 2.11 – Aplanatic objective consisting of two components.

index as that of the first lens ($n_{D1} = 1.84666$), the ray originating in object point A travels to point B with no change in direction and $S_2 = (-t_0 + r_2)$. On the other hand, if point A is the aplanatic point of the second (spherical) surface one may use Eq. (2.26) and to get

$$S_2 = \frac{n'_2 + n_2}{n_2} r_2 = r_2 - t_0 \tag{A}$$

which gives

$$r_2 = -t_0 \frac{n_2}{n_2'} = -0.7 \times \frac{1.84666}{1} = -1.29 \text{ mm.}$$

Then $S_2 = -1.29 - 0.7 = -1.99$ mm and Eq. (2.26) yields

$$S'_2 = S_2 \frac{n_2}{n'_2} = -1.99 \times 1.84666 = -3.67 \text{ mm.}$$

Magnification of the first component is governed by Eq. (2.27): $V_1 = 1.84666^2 = 3.41$. Then the magnification of the second lens should be $V_{t2} = 6/3.41 = 1.76$. According to the schematics of Fig. 2.42, point A' which is the image of A created by the first lens serves as a center of curvature of the third (spherical) surface (e.g., the first surface of the second component). This point is also aplanatic, but there is no bending of rays here and the magnification is determined from Eq. (2.28): $V_3 = 1/n_{D2}$. The ray travels further to point C of the fourth surface for which A' is

the aplanatic point and the image after refraction is point A". Here again the magnification is determined by Eq. (2.27): $V_4 = (n_{D2})^2$. Thus the entire magnification of the second element becomes $V_{t2} = V_3 \times V_4 = (n_{D2})^2/n_{D2} = n_{D2}$. As we see above, this value should be 1.76. The closest glass from the data of Appendix 2 is SF-11 with $n_D = 1.78472$. We choose this for the second element of the objective.

Going back to the radii of the second lens, for the third surface we have $r_3 = S'_2 - d_2 = -3.67 - 1 = -4.67$ mm. Keeping in mind that $S_4 = r_3 - d_3 = -4.67 - 3 = -7.67$ mm we get, again using Eq. (2.26)

$$r_4 = -7.67 \frac{1.78472}{1.78472 + 1} = -4.912 \text{ mm}$$

and $S'_4 = 1.78472 \times S_4 = -13.69$ mm. Thus, the objective is a compound of two elements performing imaging around the aplanatic points only. The total magnification is $V_{\text{tot}} = V_1 \times V_{t2} = 3.41 \times 1.78472 = 6.085$, i.e., 1.4% deviation from the required value (such a tolerance is usually acceptable).

It should be mentioned that the design presented here is for demonstration and teaching purposes only. In reality many more ray tracing operations followed by image quality analysis are required.

2.12. We start with the positioning of the system components and use the paraxial formulas for thin lenses. As total magnification V = 5/0.25 = 20 and $V_1 = -5$, we get

$$S'_1 = f'(1 - V_1) = 15 \times 6 = 90 \text{ mm};$$
 $V_2 = \frac{V}{V_1} = \frac{20}{(-5)} = -4$
 $S_2 = 15\frac{1+4}{-4} = -18.75 \text{ mm};$ $S'_2 = S_2 \times V_2 = 75 \text{ mm}.$

The reticle of thickness d = 2 mm makes the optical path longer by d(1 - 1/n) = 0.67 mm (see Problem 1.6). This causes the defocusing aberration in the plane P, as depicted in Fig. 2.43. Assuming the active size of the first lens is dictated by its diameter, we have for the coordinate of the plot in the plane P

$$\delta s'_{1,t} = -\frac{y_1}{S'_1} \delta_{\mathbf{R}} = -\frac{0.67}{90} y_1; \quad \delta s'_{1,t \max} = -\frac{0.67}{90} 4 = -29.6 \,\mu\text{m}.$$

Transferring this aberration to the CCD plane, we obtain the corresponding straight line on the plot, with the maximum deviation $\delta s'_{2,t \max} = \delta s'_{1,t \max} V = 29.6 \times (-4) = 118.4 \,\mu\text{m}$. This means there is a noticeable defocusing on the CCD. Correction can be done by displacement of the CCD to a new position where the aberration has the same value, but with opposite sign (shown by the dotted line in Fig. 2.43b). To calculate the displacement required to get



FIGURE 2.43 Problem 2.12 - (a) Defocusing caused by reticle and (b) the aberration plot in the CCD plane.

the defocusing of $(-118.6) \ \mu m$ we first find the active size of the second lens: $h_2 = h_1 S_2 / (-S'_1) = 0.83 \text{ mm}$, which gives the necessary displacement as follows:

$$\delta s'_2 = x = 0.1184 \times 75/0.83 = 10.84$$
 mm.

Instead of referring to the transverse aberration we could consider the lateral aberration. In such a case we get $x = 0.67 \times 4^2 = 10.8$ mm which is actually the same result.

Of course, the case considered in the problem is quite trivial, but it demonstrates the principle of aberration transfer and summation.

2.13. Considering the on-axis point we find the angular range of rays creating the image as $\pm \beta = \pm \arctan(D/2S') = \pm \arctan(1/2f) = \pm 23^\circ$ (it is assumed here that the difference between *S'* and focal length is small while imaging distant objects). As is evident from the aberration plot, the residual (uncorrected) aberrations of lens L are significant and cannot be neglected. Correction by additional elements of the system is desirable. This can be realized in a layout with a pentaprism and cannot be done if a mirror is used for bending. The prism introduces additional spherical aberration described by Eq. (A) of Problem 2.2. By substituting in that expression different values of *u*, from 0° to 23°, we obtain the plot shown in Fig. 2.44a by the solid line. Comparing this to the residual aberration of the lens, shown by the dotted line, we find that they have opposite sign and therefore noticeable correction for the maximum angle $u = 23^\circ$ where aberration of the lens is as high as 1.1 mm. Since the penta-prism is equivalent to a parallel glass slab of thickness $t_e = 3.414a$, where *a* is the entrance face size (see Section 1.4)



FIGURE 2.44 Problem 2.13 – an aberration plot: (a) for separate elements (1, lens; 2, penta-prism); (b) residual values after correction.

for unfolded diagram of the prisms), we get

$$\frac{3.414a}{n} \left(1 - \frac{n \cos u_{\max}}{\sqrt{n^2 - \sin^2 u_{\max}}} \right) = 3.414a \times 0.031 = 1.1 \text{ mm}; \quad a = 10 \text{ mm}.$$

Calculating the aberration of the prism for different angles and subtracting the results from the plot of the lens aberration we obtain the plot of the final residual aberrations of the system after correction (Fig. 2.44b).

2.14. In general it can be stated that a symmetrical configuration with a parallel trace between two components yields the best results with regard to residual aberrations. In our case each lens is optimized for imaging from infinity to its focal plane. Then, according to the rules described in Section 2.1.6 we should find the aberrations of lens L_1 in reverse operation mode (light propagating from the right to the left) and transfer them, keeping in mind the linear magnification of the system, to the focal plane of lens L_2 where they are added to the aberrations of the second lens. Obviously the smaller the aberration of separate elements the lower the total sum of aberrations in the bundle entrance.

Optimization of the first lens can be done using Eqs. (2.15) and (2.16). For $n_D = 1.5168$ we get the radii of the lens as follows (r_1 is going towards the

parallel beam, i.e., inside the condenser; h = D/2 = 15 mm):

$$\xi = 1/1.5168 = 0.6593;$$
 $r_1 = -2 \times (1 - 0.6593) \frac{2.319}{2.6593} 60 = -35.65 \text{ mm};$
 $r_2 = 2 \frac{0.3407 \times 2.319}{2 - 0.6593 - 1.739} = 238.2 \text{ mm}.$

Similar calculations for the second lens give

$$r_3 = -r_1 \frac{f'_2}{f'_1} = 71.3 \text{ mm}; \quad r_4 = -r_2 \frac{f'_2}{f'_1} = -238.2 \times 2 = -476.4 \text{ mm}.$$

Now, using Eq. (2.15) we find the transverse aberrations of separate components:

$$\delta s'_{t1} = -\frac{1}{8} \frac{0.6593 \times 3.3407}{0.1161 \times 2.319} \frac{h^3}{f'_1^2} = -0.959 \text{ mm};$$

$$\delta s'_{t2} = -\frac{1}{8} \frac{0.6593 \times 3.3407}{0.1161 \times 2.319} \frac{h^3}{f'_2^2} = -0.24 \text{ mm}$$

and the total transverse aberration in the plane of the bundle entrance:

$$\delta s'_{\text{tot},t} = -\delta s'_{t1}V + \delta s'_{t2} = 0.959 \times 2 - 0.24 = 1.68 \text{ mm}.$$

It is interesting to mention that if two components were identical $(f'_1 = f''_2)$ then $\delta s'_{t1} = \delta s'_{t2}$; |V| = 1; and the total transverse aberrations of the condenser would approach zero.

2.15. Expression (2.33) defines the minimum resolvable spot as follows:

$$\delta_{\rm dif} = AB = \frac{1.22\lambda p}{D_{\rm p}} = 0.5 \ \mu \rm{m}.$$

Assuming the entrance pupil is located at the mounting of the imaging lens we have p = 30 mm. Then, taking also $\lambda = 0.5 \,\mu$ m, we get the necessary size of the lens (the entrance pupil) as

$$D_{\rm p} = \frac{1.22 \times 0.5 \times 30}{0.5} = 36.6 \,\,{\rm mm}.$$

2.16. We use Eq. (2.36), keeping in mind that the illumination angle cannot be greater than that defined by the numerical aperture of the objective. Therefore for the first two objectives only a fraction of NA_C can be exploited and the resolution is

$$R_1 = \frac{\lambda}{2\text{NA}} = \frac{0.5}{2 \times 0.25} = 1 \,\mu\text{m}; \quad R_2 = \frac{\lambda}{2\text{NA}} = \frac{0.5}{2 \times 0.65} = 0.38 \,\mu\text{m}.$$

The NA of the last objective is greater than that of the condenser and in this case from Eq. (2.36) we get

$$R_3 = \frac{\lambda}{\text{NA} + \text{NA}_{\text{C}}} = \frac{0.5}{1.2 + 0.96} = 0.23 \,\mu\text{m}.$$

2.17. The key equation here is

$$\sin\beta - \sin\beta_0 = \frac{m\lambda}{d} \tag{A}$$

where β_0 is the illumination angle and β is the angle of the diffraction maximum of order *m*. If m = -1 and $\beta = -\beta_0$ we have the situation depicted in Fig. 2.25b and the resolution is improved by a factor of 2. This occurs if the following condition is obeyed:

$$\tan \beta_0 = \frac{D/2}{f'} = \frac{2.5}{16} = 0.156; \quad \beta_0 = 8.9^\circ.$$

Expression (2.35) renders resolution in this case: $d = 0.5/2 \sin 8.9^\circ = 1.6 \,\mu\text{m}$. If the illumination angle is smaller than 8.9° we still can improve the resolution, as follows from the geometry shown in Fig. 2.45. In this case the position of the zero order is determined by $y_0 = f' \tan \beta_0 = 16 \times \tan 5^\circ = 1.4$ mm whereas the (1)st order comes at the side of the aperture stop:

$$\tan \beta = 2.5/16 = 0.156; \quad \beta = 8.9^{\circ}.$$

Then, from Eq. (A) one obtains the resolution as follows:

$$d = \frac{0.5}{\sin 5^\circ + \sin 8.9^\circ} = 2.07 \,\mu\text{m}.$$

2.18. For a diffraction-limited system the parameter z in Eq. (2.37) becomes

$$z = \frac{\pi Dd}{2\lambda p} = 3.8317 \frac{d}{d_{\rm dif}}$$

where d_{dif} , determined from Eq. (2.32), is almost the full spot size. For a circle of $d = 0.5 d_{\text{dif}}$ we get z = 1.91 and Eq. (2.37) yields L = 60%. This means that if our system is limited by diffraction only up to 60% of the full energy



FIGURE 2.45 Problem 2.17 - (a) Oblique illumination in a microscope objective and (b) location of diffraction maxima in the aperture stop.

of the spot will be collected inside 50% of its diameter. Since this differs from what was actually found (45%), one can draw the conclusion that the system has noticeable aberrations.

2.19. The maximum spatial frequency transferred by the system (the cut-off frequency) is determined from Eq. (2.45):

$$v_{\rm C} = \frac{2{\rm NA}}{\lambda} = \frac{2 \times 0.25}{0.6 \times 10^{-3}} = 833$$
 cycles/mm

which corresponds to a period of 1.2 μ m in the object space or 12.0 μ m in the CCD plane (after ×10 magnification). According to the Nyquist theorem the sampling frequency has to be two times higher than the tested frequency, i.e., two pixels of the CCD are necessary for one period of 12 μ m. Therefore the CCD elements (pixels) should be arranged with a 6 μ m pitch (center-to-center distance).

2.20. We start with the calculation of modulation, M_0 , in the object space. Let the illumination intensity be I_0 and it be spread uniformly on the target. The square wave segments of the target coated with chrome reflect 0. $7I_0$ whereas the target segments with no coating reflect 0. $04I_0$. This gives the following value for the modulation of the object:

$$M_0 = \frac{0.7I_0 - 0.04I_0}{0.7I_0 + 0.04I_0} = 0.89.$$

We denote the background scattered light as I_S and assume that it is the same at all locations (uniformly spread in the system). Therefore the true modulation in the image plane, M_i , and the apparent modulation registered with the scattered light, M'_i , are related as follows:

$$\frac{1}{M'_{i}} = \frac{I_{\max} + I_{\min} + 2I_{S}}{I_{\max} - I_{\min}} = \frac{1}{M_{i}} + \frac{1}{M_{S}} = \frac{1}{\text{MTF}' \times M_{0}}$$
(A)

where MTF' is the measured MTF value when the scattering is present. The influence of the scattered light, $1/M_S$, can be found from Eq. (A) and we will find it separately for low frequencies and for high frequencies. Since scattering does not depend on the frequency chosen for measurement we can write:

$$\frac{1}{M_{\rm S}} = \frac{1}{M'_{\rm iH}} - \frac{1}{M_{\rm iH}} = \frac{1}{M'_{\rm iL}} - \frac{1}{M_{\rm iL}};$$
$$\frac{1}{M_{\rm iH}} = \frac{1}{M'_{\rm iH}} - \frac{1}{M'_{\rm iL}} + \frac{1}{M_{\rm iL}} = \frac{1}{M'_{\rm iH}} - \frac{1}{M'_{\rm iL}} + \frac{1}{M_0}$$
(B)

where it is taken into account that at very low spatial frequency MTF is close to 100% (if no scattering is present in the system): $M_{iL} = M_0$. Using in Eq. (B) the

definition of MTF as per Eq. (2.43), we get for the true MTF at high frequency

$$MTF_{H} = \frac{M_{iH}}{M_{0}} = \left(\frac{1}{MTF'_{H}} - \frac{1}{MTF'_{L}} + 1\right)^{-1}$$
(C)

and by substituting the problem data in Eq. (C) we have

$$MTF_{\rm H} = \left(\frac{1}{0.2} - \frac{1}{0.7} + 1\right)^{-1} = 0.22.$$

Obviously, if there is no scattering MTF'_L approaches unity and $MTF_H = MTF'_H$.

2.21. We accept that the minimum overall MTF of the system, optics + CCD + monitor, should be 5% at least for a spatial frequency which can be observed. Since the total MTF is the product of the MTFs of separate elements, we obtain the minimum requirements for the MTF of the imaging optics: $MTF_o = MTF_{tot}/MTF_{CCD} = 0.05/0.6 = 0.083$. As our system is diffraction limited its MTF obeys Eq. (2.44) with cut-off frequency (Eq. (2.45)) given by

$$v_{\rm C} = \frac{2{\rm NA}}{\lambda} = \frac{2 \times 0.15}{0.5 \times 10^{-3}} = 600$$
 cycles/mm

where MTF = 0. Although the graph of MTF(v) described by Eq. (2.44) is a curve, its deviation from a straight line is not very noticeable. Then, approximating the graph by a straight line we find the frequency which corresponds to MTF = 0.083. This frequency is v = 600(1 - 0.083) = 550 cycles/mm. Therefore, it is impossible to see on the monitor the details originating in a spatial frequency of 575 cycles/mm in the object plane.

2.22. We choose the architecture of the telecentric configuration as that of Fig. 2.30b and restrict ourselves to the paraxial range. Then we have

$$f'_1 + f'_2 = 100 \text{ mm}; \quad f'_2 = f'_1 \times |V| = 3f'_1; \quad f'_1 = 25 \text{ mm}; \quad f'_2 = 75 \text{ mm}.$$

To find the size of the aperture stop we should calculate the required NA in the object space. Suppose the system is free of aberration. Then, using the first relation of Eq. (2.35) and keeping in mind that the required resolution should be equal to $2 \mu m$, we obtain

$$d = \frac{\lambda}{\text{NA}} = 2 \,\mu\text{m};$$
 NA = 0.5/2 = 0.25 = sin $u_{\text{max}};$ $u_{\text{max}} = 14.5^{\circ}.$

This yields the size of the aperture stop as follows (see Fig. 2.46): $D_{ab} = 2f'_1 \tan u_{\text{max}} = 2 \times 0.258 \times 25 = 12.9 \text{ mm}$. By considering further the angular field of view and taking into account that the marginal chief ray should pass through the center of the aperture stop, we get $y = f'_1 \tan \beta = 25 \times \tan 5^\circ = 2.2 \text{ mm}$ and



FIGURE 2.46 Problem 2.22 – A telecentric system.

the size of the first lens becomes $D_1 = 2h_1 = 2(f'_1 \tan \beta + D_{ab}/2) = 2(25 \tan 5^\circ + 12.9/2) = 17.3$ mm. The size of the second lens is calculated in a similar way: $D_2 = 2h_2 = 2(f'_2 \tan \beta + D_{ab}/2) = 26$ mm. The object is positioned 25 mm in front of lens L₁ and the image is created at a distance of 75 mm behind lens L₂.

2.23. Referring to the system depicted in Fig. 2.31, we use the second approach and Eq. (2.49) which gives for the configuration with a distance l = 60 mm between the first lens and the image

$$f'_{\rm e} = 60/0.75 = 80$$
 mm; $f'_1 = -f'_2 = 40$ mm; $S'_F = 40$ mm

and therefore the distance between the lenses is 20 mm. Obviously Petzval's sum is equal to zero and therefore the image curvature is negligible.

Sources of Light and Illumination Systems

3.1. Thermal Radiation Sources for Visible and IR

Thermal radiation sources, like filament lamps or Nernst rods, have been exploited for many years and are still in use in many optical systems, primarily in those intended for imaging. The modern quartz tungsten halogen (QTH) lamps and IR emitters are just technologically improved versions of the older sources.

The operation of these sources is based on thermal radiation laws described in detail in Chapter 6. We will address here some specific features of thermal sources that affect their use in practice.

A QTH lamp has an electrically heated tungsten filament positioned inside a transparent bubble made of fused silica and filled with halogen gas. This gas causes a chemical reaction between the tungsten atoms evaporated from the filament and deposited on the bubble wall and the halogen molecules improving in such a manner both the lifetime of the lamp and the transparency of the QTH envelope.

A QTH lamp is a source of broadband radiation: actually the tungsten itself emits at all wavelengths, but the transparency of the envelope limits the useful emission to visible and near-IR wavelengths (up to about 2.5 μ m). Actually some UV radiation is also available, in the wavelength interval from 240 to 400 nm, although this is of low intensity.

Usually the "optical strength" of the light source is characterized by its irradiance, E_{λ} , measured as the radiation flux, per 1 nm wavelength band, incident on an area of 1 m² at a distance of 0.5 m from the source. The second important feature of


FIGURE 3.1 Spectral irradiance of QTH lamps (*The Book of Photon Tools*, Oriel Instruments (2003), with permission of Spectra Physics Ltd).

the lamp is its color temperature, T_c (see definition in Section 6.2). There exist QTH lamps with color temperatures of 2,850 K and of 3,200 K (even up to 3,400 K). Although the emissivity of tungsten is strongly selective, it increases rapidly in the visible where it achieves a value of 0.8, providing continuous radiation which is close to that of a black body. A typical graph of irradiance of QTH lamps of 100 and 250 W is shown in Fig. 3.1. The radiated spot in a QTH lamp usually has the shape of a rectangular (lower power) or a coiled filament (larger lamps) of several millimeters in size.

The lifetime of the lamps varies from 50 to 1,000 hours and it is evidently a critical parameter. One can improve it significantly by reducing the voltage (which is accompanied by decrease of the temperature). A voltage reduction of 6% might result in a doubling of the lifetime. However, a reduction of more than 10% becomes problematic as it could spoil the halogen cycle inside the lamp bubble.

Much more intense radiation than that of QTH lamps is produced by arc lamps where an electrical discharge arc is created in surrounding xenon, mercury, deuterium, or other inert gas. The color temperature of such lamps can be 4,000 K and even as much as 6,000 K (xenon lamp). Another important feature is a great number of spectral lines in the UV. With a deuterium arc lamp wavelengths as short as 160 nm can be obtained. The brightest portion of the arc is usually of several millimeters in size, but its location might be unstable.

If mid- or far-IR wavelengths are required then special IR emitters should be exploited. The Nernst rod made of zirconium ceramics and heated to a color temperature of about 2,000 K was one of the first wideband sources. Another kind of IR emitter in use is the silicon carbide ceramic element of 1,300 K color temperature. In both sources the radiating element is a cylinder of several millimeters in diameter heated by a DC electric current of about 4–5 A. In the wavelength range from 1 to about 28 μ m IR emitters have a smooth continuous spectrum.

Problems

3.1. A QTH lamp operated at 12 V DC has a filament of 4. 2×2.3 mm heated to a temperature of 3,234 K. Assuming a tungsten emissivity of 0.8, find the spectral irradiance for a wavelength of 0.5 μ m: (a) at nominal voltage; (b) after the voltage is reduced by 5%.

[Note: For the given source the temperature and the voltage supplied are related to each other as follows: $d(T/T_0)/d(U/U_0) = 0.4$.]

3.2. The lamp described in Problem 3.1 is used as the source of a linear illumination system. At a distance of 60 mm from the lamp filament a plane convex cylindrical lens is positioned. The lens is made of BK-7 glass, its refraction surface is of 20.65 mm radius and its size is 20 mm (height) by 100 mm (width). Find the location of the illuminated line and the intensity distribution along it.

3.2. Lens-based Illumination Systems

In a variety of optical architectures illumination is generated inside the system by a module or sub-assembly, which is an integral part of the whole configuration. We will consider such a module as a separate illumination system.

All illumination systems are intended either for the creation of stratified light (a pattern of a special shape, like a straight line, or a ring, or a more complex form) or for illuminating an object in an imaging arrangement. Examples of systems from the first group are considered in Problems 3.2 and 3.12. In this section we discuss illumination for imaging optics.

In general, uniform illumination of the field of view is of our main concern. To achieve this goal the principal rule is to avoid the creation of the light source image in the object plane or in the image plane (and also not in the vicinity of these planes). There exist several ways to do this. An illumination system with a single lens is shown in Fig. 3.2. The lens L transfers the image of the light source S into the entrance pupil of the imaging optics (objective L_{ob} in this case) which builds



FIGURE 3.2 Single-lens illumination system.



FIGURE 3.3 Two-lens illumination system.

the image of the object y in the plane y'. Illumination on lens L is of the highest uniformity since each part of the source S contributes light to each point in the plane L (in the figure the rays originating in the center of the source are drawn as solid lines whereas the dotted lines are related to points A and B at both sides of the filament). The object is positioned very close to lens L and therefore it is not affected by non-uniformity of the source S.

The configuration depicted in Fig. 3.3 comprises two lenses, L_1 and L_2 , for illuminating the object y. The source image is transferred by the first lens into the plane of the second lens where a diaphragm of variable size is positioned. Changing the diaphragm enables one to select illumination from a different part of the source. Lens L_2 builds the image of lens L_1 in the object plane y. Again, the highest uniformity is achieved here because all points of the source S contribute radiation to each point of the object. The drawback of the configuration becomes evident if we consider the side rays coming to the imaging objective L_{ob} : some rays originating in the source S are cut by the final size of the objective lens which might result in considerable vignetting. This problem is eliminated in the three-lens architecture shown in Fig. 3.4. The first two lenses and the source S are located and function as in the previous case of the two-lens system. An additional lens L_3 transfers the image of S further into the plane of the entrance pupil P of the imaging optics objective, L_{ob} , providing in such a way that all relevant rays from



FIGURE 3.4 Three-lens illumination system.



FIGURE 3.5 Illumination system of a microscope.

the source, either from the on-axis point or from the filament sides, participate in the creation of the image in the plane y'. The object y should be positioned very close to lens L_3 where illumination is uniform.

The principles explained above are implemented in the microscope illumination module depicted in Fig. 3.5, which is applicable in the situations where the observed object, y, is not transparent (opaque illuminator). This frequently occurs in metallography, the semiconductor industry, and other important applications. The microscope objective, L_{ob} , creates the image of y in the plane y' at magnification V and the cubic beam splitter, BS, is introduced in order to provide illumination of the object from above, through the same objective lens. The aperture stop D₁ limits the size of the light source S exploited for illumination and its image is transferred by lenses L₁ and L₂ into the aperture diaphragm D₃ of the microscope. The stop D₂ serves as a field diaphragm: its location is conjugate with the object plane y. By varying the size of D₂ one can choose the field of view under illumination.

Problems

3.3. In the arrangement shown in Fig. 3.3 an object y of 10 mm in size is illuminated by a radiation source S of 3 mm by 3 mm followed by two identical lenses of 30 mm in diameter. The total length of the arrangement is fixed as l = 250 mm (assuming the thickness of the lenses can be neglected).

- (a) What should be the optical power of the lenses and where should they be located in order to ensure maximum illumination level and maximum uniformity of illumination at y?
- (b) What is the maximum useful size of the source S in this configuration?

3.4. In the microscope illumination system shown in Fig. 3.5 the source S is a fiber bundle of 6 mm diameter and numerical aperture NA = 0.25. The bundle is positioned in the focal plane of lens L_1 . Lens L_2 of 50 mm focal length is followed by the beam splitter installed in the imaging branch of the instrument, where the objective L_{ob} of 40 mm focal length projects the magnified image (linear magnification V = -2) of the object y onto an area sensor of 4 mm by 4 mm. Stop D_3 of 12 mm diameter acts as the aperture diaphragm of the imaging optics. Find:

- (a) the optimal size of the field diaphragm D₂ positioned in the middle of the distance between L₂ and L₃;
- (b) the minimum diameter of each lens of the illumination system.

3.5. In a luminescent microscope operated with an objective of $\times 10$ magnification and NA = 0. 185 illumination is carried out by a mercury lamp of 28 mW/m²/nm irradiance at a 240 nm wavelength followed by the three-lens configuration shown in Fig. 3.4. The field of view (FOV) of the objective is 1.8 mm and its entrance pupil P is positioned 18 mm from the observed object. Lens L₁ is of 10 mm in size and is of 13 mm focal length. The active spot of the lamp a = 1.0 mm.

- (a) Find the parameters of the illumination system (sizes, focal lengths, location of the elements).
- (b) Calculate the number of photons per second incident on a cell of 6 μ m in size located in the FOV of the microscope.

3.3. Lasers

3.3.1. Main Characteristics of a Laser Beam

Lasers differ significantly from all other sources of radiation, primarily due to the fact that lasers generate stimulated radiation in strongly non-equilibrium



FIGURE 3.6 Schematic of laser light generation.

conditions, in contrast to thermal radiation sources, for example, which emit spontaneous radiation in a thermal equilibrium state (or close to it). As a result, laser light (i) is highly coherent; (ii) is highly monochromatic; (iii) propagates as a highly parallel beam (very small divergent angles); and (iv) usually has a well-defined polarization.

How do lasers work? There are many books devoted to these fascinating light sources which have become so popular in the last 40 years (e.g., see Young, 1984; Yariv, 1982). We have no intention of discussing here the details of laser design and operation and will concentrate only on those features of laser radiation which are important for applications.

In practice a laser is configured as a Fabry–Perot etalon (see Section 5.4) where the spacing between two mirrors (M1 and M2 in Fig. 3.6) is filled by an optically active material – a medium in which a reverse population of excited atoms can be achieved for a relatively long period of time. Starting with a single photon emitted spontaneously in the direction normal to the mirrors an avalanche of secondary photons is generated (stimulated radiation). Most of the photons are reflected by the mirrors back to the laser cavity, but a portion is transferred through the mirror out of the device and it is this portion which constitutes the beam of light emitted by the laser. The basic idea of laser light generation is depicted in Fig. 3.6.

A great variety of lasers are available for optical use. They cover the spectral interval from UV to mid-IR, can be operated either in continuous or in pulse mode (duration of each pulse varies from microseconds to femtoseconds), and can supply optical power from several microwatts to megawatts. The schematic shown in Fig. 3.7 relates to all these types of lasers.

We consider a laser cavity (sometimes referred to as an optical resonator) of length L with two identical plane mirrors of high reflectivity R and negligible absorptivity. The resonator is assumed to be axially symmetric with the optical axis OZ centered at the middle of the cavity. It can be shown that the electric field



FIGURE 3.7 (a) Laser cavity and laser beam parameters and (b) radial profile of the laser light intensity.

of the electromagnetic wave generated in the resonator and propagating outside is described by the following expression:

$$E(r,z) = E_0 \left\{ \frac{w_0}{w(z)} \exp\left[-\frac{r^2}{w^2(z)}\right] \right\} \times \exp\left\{-i\left[kz - \tan^{-1}\left(\frac{z}{z_0}\right)\right] \right\}$$
$$\times \exp\left[-i\frac{kr^2}{\rho(z)}\right]$$
(3.1)

where

$$z_0 = \frac{\pi w_0^2}{\lambda}; \quad w^2(z) = w_0^2 \left[1 + \left(\frac{\lambda z}{\pi w_0^2}\right)^2 \right]; \quad \rho(z) = z \left[1 + \left(\frac{\pi w_0^2}{\lambda z}\right)^2 \right]$$

and $k = 2\pi/\lambda$ is the wavenumber. Of the three terms on the right-hand side of Eq. (3.1) only the first (called the amplitude phase) is relevant for calculation of light intensity, *I*, as the others (the longitudinal phase and the radial phase) are eliminated while computing the real values:

$$I(r,z) = E \times E^* = |E|^2 = I_0 \exp(-ar^2); \quad \left[a = \frac{2}{w^2(z)}\right]. \tag{3.2}$$

Hence, the light beam is not of constant intensity, but has a radial profile of a Gaussian function at any cross-section perpendicular to the optical axis (for each chosen coordinate z). The Gaussian is overspread to infinity (in the radial direction), but, in order to define anyhow the beam radius, it is usually accepted to address only the part of the beam where the light amplitude is reduced as 1/e, e.g., intensity is reduced to $1/e^2 = 0$. 135 of the on-axis value. The corresponding radial coordinate is w(z) (see Fig. 3.7b) and in the middle of the resonator, at z = 0, it is denoted as w_0 . This is the smallest achievable value for a given resonator and this cross-section is defined as the laser beam waist. The parameter $\rho(z)$ from Eq. (3.1)

represents the curvature of the wave front at coordinate z. Obviously at the beam waist (z = 0) the curvature is equal to infinity.

Connecting all points of w(z) we get two curves shown by the dotted lines in Fig. 3.7a. Calculating the derivative dw/dz, using Eq. (3.1), we find the slope of the graph, or actually two straight lines passing through the center of the coordinate system, z = 0, to which the curves asymptotically approach (full lines on the figure). The angle, 2θ , between these two lines defines the divergence of the laser beam:

$$2\theta = 2\frac{\mathrm{d}w}{\mathrm{d}z} = 2\frac{\lambda}{\pi w_0}.\tag{3.3}$$

The waist radius, w_0 , is related to the resonator size and the wavelength as

$$w_0 = \sqrt{\frac{\lambda L}{2\pi}}.\tag{3.4}$$

Regarding the beam size at the exit from the cavity, it can be shown that $w(L/2) = \sqrt{2} \times w_0$.

Returning to the intensity distribution of Eq. (3.2) we should define the value I_0 . We express it in terms of the full optical power of the beam, P, which remains the same at any cross-section z. By integrating I(r, z) along the radius up to infinity, at a given z we obtain

$$P = 2\pi I_0 \int_0^\infty \exp(-ar^2) r dr = \frac{\pi I_0}{a}$$
(3.5)

and therefore

$$I_0(z) = \frac{2P}{\pi w^2(z)}.$$
(3.6)

The parameters defined in Eqs. (3.2)–(3.6) are the main features of a laser beam required for most applications.

As to the polarization of a laser beam, it should be mentioned that, in general, radiation coming out of the cavity is linearly polarized. However, although each avalanche of photons at any short time interval is constituted from photons of the same polarization, it might vary randomly if a longer period of time is involved. Special measures should be undertaken to preserve the polarization of the specific kind required (usually linear polarization in the vertical plane).



FIGURE 3.8 Problem 3.7 – A laser surveying system.

Problems

3.6. A He–Ne laser comprises a resonator of 300 mm in length which generates a continuous beam of $0.6328 \,\mu$ m wavelength. Find the beam size and the divergence angle at the exit of the laser.

3.7. A laser surveying system (see Fig. 3.8) comprises a transmitter with a He–Ne laser of 30 cm length and 2 mW power and a receiver with a detector of 1.5 mm in size, a dynamic range of 10^4 , and a minimum detectable power (NEP) of 10^{-7} W. What is the minimum and maximum working range of the system?

3.3.2. Beam Expansion and Spatial Filtering

Beam expansion and spatial filtering are related to operation with additional optical elements attached to a laser. Since laser beams have a Gaussian shape, as explained above, and propagation of such a beam through a system of lenses has some specific features, we consider first the rules governing the propagation of a Gaussian beam. The beam waist in front of the lens has a size $2w_1$ and after propagation through the lens it is $2w_2$ (see Fig. 3.9). The distances *S* and *S'* between the lens and the



FIGURE 3.9 Propagation of a Gaussian beam through a lens.



FIGURE 3.10 Configuration of a beam expander.

beam waists are related to each other as follows:

$$\frac{1}{S'} - \frac{1}{S} \times \frac{1}{1 + \frac{z_{\rm R}^2}{S(S+f')}} = \frac{1}{f'}$$
(3.7)

where $z_{\rm R} = \pi w_0^2 / \lambda$ and $w_2 = w_1(S'/S)$.

Beam expansion becomes important in situations where (i) the laser beam is to be transmitted over a large distance, as in optical communication systems, for instance; or (ii) the laser light is to be concentrated in a very small spot, as in material processing procedures. In the first case the divergence angle of the laser beam has to be reduced significantly in order to keep the light energy concentrated in a small size spot even after traveling a distance of hundreds or thousands of meters. In the second case the beam diameter should be increased drastically to reduce as much as possible the diffraction limit of the lens concentrator. Both goals can be achieved by inserting an inverted telescope just after the laser. An example is shown schematically in Fig. 3.10. Radiation coming from the laser is collected by the first lens, L₁, in the vicinity of the mutual focus of both lenses of the telescope and then proceeds to the second lens, L₂, where it is defocused. Remembering that the angular magnification of a telescopic system is defined as $W = f'_1/f'_2$ (see Eq. (1.18)), we have

$$\beta' = \beta W$$
; and $D' = D/W$ (3.8)

and the diameter of the laser beam after traveling the distance l is

$$\tilde{D} = D/W + l\beta W. \tag{3.9}$$

The configuration shown in Fig. 3.10 sometimes might cause problems in practical applications. For example, if a high-power laser beam is traveling through the beam expander of Fig. 3.10 all the optical power is inevitably concentrated in the vicinity of the focal plane of the lenses, inside the expander. The energy density here might become very high, even to a level capable of destroying the system elements. To avoid such an undesirable situation a beam expander with a Galilean telescope (see Section 1.2) is exploited, with the negative lens towards the laser, so that the focus becomes a virtual point and there is no dangerous concentration of energy inside the system.

Spatial filtering is actually the procedure of "cleaning" a laser beam. To explain what can be "cleaned out" consider again the basic laser cavity, keeping in mind that it acts as a resonator and like any other resonator it can be characterized by eigen values and eigen functions. These eigen functions are oscillations which can be self-generated, in a sporadic manner, by the resonator. As far as an optical resonator is concerned such eigen functions are different modes of electromagnetic waves developed in the laser cavity. Due to the vectorial nature of electromagnetic fields these modes constitute a two-dimensional array of functions usually indexed like a tensor, T_{ij} , where i, j = 0; 1; 2; ... The first mode, called TEM₀₀ (transverse electromagnetic mode of zero-zero order), is the basic one and it is this mode that has a radial distribution of intensity described by the Gaussian function (Eq. (3.2)), with the maximum energy density on the optical axis. Other possible modes, like TEM₁₀, TEM₀₁, TEM₁₁, and others, have an intensity distribution which differs from Eq. (3.2) – they might have several points of maximum intensity (in each cross-section perpendicular to OZ), or points of maximum intensity arranged as a ring shape, and so on. To be accurate, we should also mention that as well as the TEM modes in the resonator there are also longitudinal modes (for further details, see Yariv, 1982).

The generation of higher-order modes might be caused by impurities in the laser cavity or just by particles on the laser resonator mirrors or by many other causes emerging unpredictably in the laser. All of them cause random variations (fluctuations) in the laser beam intensity, sometimes called spatial noise. There are numerous applications where the spatial noise and the higher-orders modes are not desirable and have to be eliminated ("cleaned out") from the laser beam. This is done by spatial filtering (see Fig. 3.11). A spatial filter is actually a lens followed by a very small pinhole positioned in the location of the beam waist (which is close, as we remember, to the focus of the lens). The size of the pinhole, d_{sf} , is dictated by the properties of the beam incident on the lens (if the lens is large enough and does not truncate the Gaussian profile of the beam):

$$d_{\rm sf} = c2\theta f' \tag{3.10}$$



FIGURE 3.11 Configuration of (a) a spatial filter and a laser beam profile (b) before and (c) after spatial filtering.

where f' is the lens focal length, 2θ is the beam divergence at the entrance to the filter, and c > 1 is a factor introduced in order to be on the safe side in eliminating truncation of the beam (usually the recommended factor is c = 1.3-1.5). To be more accurate, it is necessary to replace f' in Eq. (3.10) by the value S' as in Eq. (3.7) describing correctly the propagation of a Gaussian beam, but in practice the approximation of Eq. (3.10) is good enough for most spatial filters.

Problems

3.8. An optical set-up (Fig. 3.12) includes a He–Ne laser (wavelength 0.63 μ m, cavity size 300 mm) and a lens of 80 mm focal length positioned first at a distance $a_1 = 5$ mm from the laser exit and then moved along the optical axis to a distance $a_2 = 100$ mm. Calculate:

- (a) the distance which moves the beam waist beyond the lens;
- (b) the maximum achievable distance between the lens and the beam waist.

3.9. A laboratory laser-guided robot (see Fig. 3.13) comprises a He–Ne laser of 5 mW power and 500 mm length followed by a beam expander with two lenses



FIGURE 3.12 Problem 3.8 – A laser followed by a lens.



FIGURE 3.13 Problem 3.9 – Schematic of a laser-guided robot.

 $(f'_1 = 8 \text{ mm}; f'_2 = 80 \text{ mm})$. On the detector side there is a lens of 10 mm in diameter. Calculate the optical power registered by the robot detector when it is 3 m from the laser source.

3.10. An argon laser generates a light beam of 514 nm wavelength and with beam waist radius $w_0 = 0.7$ mm. This is followed by a beam expander built of two lenses: L₁, diameter = 4 mm, $f'_1 = 8$ mm; and L₂, diameter = 40 mm, $f'_2 = 100$ mm.

- (a) Find the beam divergence in front of the beam expander and behind it.
- (b) For cleaning the beam a spatial filter comprising a lens L₃ (10 mm diameter, 20 mm focal length) and a pinhole is added to the system. Calculate the pinhole size for two cases: when the spatial filter is positioned after the laser, in front of the beam expander; and when the filter is positioned just after the beam expander.
- (c) Find the relative amount of energy of the laser passing through the pinhole in the second case of (b).

3.11. On a construction area a laser transmitter transfers a reference beam to a distance of S = 200 m. The laser is operated at an IR wavelength of 0.83 µm and the divergence at the exit of the transmitter sub-assembly M (see Fig. 3.14) is



FIGURE 3.14 Problem 3.11 – Schematic of a laser reference system for a construction area.

 $2\theta = 10^{-3}$ rad. The beam diameter here is $D_0 = 3$ mm and lens L₁ is of 10 mm focal length. On the transmitter side there is a detector of 5 mm in size. What lens should be added to lens L₁ of M in order to ensure that the beam diameter will not be greater than the detector size at all distances?

3.3.3. Laser Diodes

These are the smallest lasers commercially available at the present time. Due to very compact design and ruggedness they have been introduced in a great variety of application areas, such as reading heads of DVD players, optical pumping of high-energy lasers, communication systems, and medical uses.

From the physical point of view a laser diode is a small resonator located in a central part of a semiconductor p–n junction arrangement (see Fig. 3.15) where a DC electric current is directly transferred into coherent radiation. Most laser diodes are made of AlGaAs or other semiconductor substrates with similar optical properties. As a result, the operating wavelengths are in the near-infrared and visible (mainly red) intervals.



FIGURE 3.15 Layout of a laser diode.

A great advantage of laser diodes is the low power consumption needed for normal operation. This results from the high efficiency of direct energy transformation from electrical to optical power. An efficiency of 30% is usually achieved when the total available optical power can be as high as tens of watts (even up to kilowatts in a laser stack configuration). In addition, direct energy transformation enables one to modulate easily the output optical power, just by changing the input electric current at a desired frequency.

Laser diodes feature some properties which make them different from other lasers and some special measures are required to fit the lasers to specific applications. We briefly describe some of them here.



FIGURE 3.16 Optical layout with a laser diode.

Because of the small size of the laser cavity the divergence angle of radiation emitted by a laser diode is much greater than that of other laser sources. Divergence of up to 40° is usual. As a result, such a laser should be followed by collimating optics aimed at reducing the divergence of the laser beam drastically before transferring it for further use. Another feature to be kept in mind is astigmatism of radiation at the laser exit. As can be seen from the schematic in Fig. 3.15, the laser diode cavity is not axially symmetric and the output spot sizes in the vertical and horizontal planes are not equal. This means that the beam waists in OX and OY differ one from another and even are not located at the same point at the laser cavity (astigmatism). Consequently the divergence angles and the beam radius in vertical and in horizontal directions, still obeying Eqs. (3.3) and (3.4), are also different. Usually $2\theta_V$ is two or three times greater than $2\theta_H$ which results in an elliptical shape of the beam generated by a laser diode. To perform "circularization" of the beam radial intensity profile either a cylindrical lens or a pair of prisms (called an anamorphic prism pair) should be introduced somewhere after the laser exit. In both cases the beam size is reduced along one axis and remains unchanged in the other direction. The operation of an anamorphic prism pair is considered in more details in Problem 3.13. The general layout of optics with a laser diode as a radiation source is presented in Fig. 3.16.

Other features which should be addressed are related to the multiple electromagnetic modes generated in the cavity. The spectral output of a laser diode comprises usually a central peak accompanied by a number of smaller peaks of other (although close to each other) wavelengths. The cause of this is a number of modes, especially longitudinal modes, of relatively high intensity which are more significant in the small resonator of a laser diode than in other situations. The total optical power is evidently spread between all generated modes. There exist applications where such a multi-mode operation is acceptable, but there are many others where a single mode is required. Special measures, like diffraction gratings, index grading, and some others, are involved in single-mode laser design in order to ensure that a single spectral line (one mode) is generated while all the others are suppressed. However, the wavelength of this single spectral line can vary with temperature. This phenomenon, called frequency hopping, might affect significantly the overall performance of a system with a laser diode. Transfer between different modes (hopping) can cause a change of the output wavelength of 1–2 nm for each 5°C temperature change (typical values).

Problems

3.12. A laser diode-based system for measuring the ground profile (Fig. 3.17) comprises a light line generator and imaging optics. The line generator consists of a cylindrical lens L₁, made of a glass rod of 7 mm in diameter, which follows a 10 mW laser diode operated at $\lambda = 680$ nm and having an astigmatic divergence of $2\theta_H = 7^\circ$; $2\theta_V = 30^\circ$. The line generator is optimized in such a manner that it provides the maximum intensity and the maximum length of the line of light on the zero level of the ground (the plane K) 1,500 mm distant from the lens.

The line of light created on the ground is imaged by a spherical lens L_2 onto an area sensor (CCD) and transferred further for image processing and calculation of the profile. The angle α between the optical axes of the two branches is small enough so that its influence can be neglected.



FIGURE 3.17 Problem 3.12 – Laser diode-based system for ground profiling.

In order to improve the image contrast with regard to the surroundings, especially in bright sunlight, an interference filter F (coating reflectivity R = 95%, FWHM $\delta\lambda = 10$ nm) is introduced in the imaging branch.

- (a) Find the intensity distribution in the generated line of light.
- (b) Assuming the sun illumination at sea level is $E_S = 1,350 \text{ W/m}^2$ and the sun temperature is 6,000 K, calculate the image contrast at the center and at the side of the light line.

3.13. A laser beam generated by a laser diode followed by a collimator has an elliptical cross-section with principal diameters of 4 mm and 8 mm. Find the anamorphic prism pair capable of correcting the ellipticity of the beam.

3.4. Light Emitting Diodes (LEDs)

In general, LEDs, like laser diodes, transform electrical energy directly into optical energy. They also comprise a semiconductor p–n junction fed by a DC current, but there is no resonator and photons are emitted spontaneously generating non-coherent radiation. The wavelengths available are not only in the IR and red regions, but also in the green and blue regions. A step in their development was a combination of several semiconductor sources generating different wavelengths in a single housing to create white light radiation. Indeed, white LEDs have become widely available in the last few years.

The spectral properties of monochromatic LEDs are inferior to those of laser diodes. Usually the bandwidth of LEDs is about 30–50 nm.

LEDs are usually operated at low voltage (2-5 V) and low current (20-100 mA) and their efficiency in energy transformation is as high as in laser diodes (up to 30%).

LEDs are manufactured in two basic configurations (see Fig. 3.18) with a flat window and with the lens incorporated as a part of the casing.



FIGURE 3.18 LED with (a) a flat window and (b) a lens.



FIGURE 3.19 Angular diagrams of a LED intensity distribution: (a) LED with a front window; (b) LED with a lens.

In applications of LEDs as light sources the angular distribution of the emitted radiation is a main concern. Examples of angular diagrams are presented in Fig. 3.19 for both types of LED design. It also should be mentioned that radiation emitted by a LED suffers from low uniformity in a cross-section perpendicular to the chip. This feature is especially noticeable at small distances from the source. Setting a diffusing glass in front of the source or even grinding the LED itself allow one to obtain much more homogeneous radiation in a wide spatial angle (an example of such an approach is given in Problem 3.14).

Problems

3.14. *Dark field illumination with a single LED.* Imaging of an opaque object in dark field illumination means that the whole field of view remains black except for some details which, due to their specular reflectivity, appear as white.

In the system depicted in Fig. 3.20 lens L_1 of 12 mm diameter and 25 mm focal length performs imaging of an object P onto an area sensor (1/2" CCD, size 4.8 mm \times 6.4 mm) at magnification $V_1 = -0.25$. The working distance (WD) defined as the free space between the object and the system has to be 16 mm at least. The illumination branch of the system which provides on-axis illumination for dark field consists of a LED followed by a condenser lens L_2 (diameter 45 mm, f # = 1.0). The LED front surface was grinded until a flat diffuse area of 3 mm in diameter was created.

Aiming at the most compact architecture, find the location of all elements of the system and a minimum size of the beam splitter, BS, required for operation in the full field of view if acceptable vignetting everywhere should not exceed 50%.

3.15. *Oblique illumination with a LED array.* Providing a minimum working distance of 16 mm, how does one incorporate in the system in Problem 3.14 a LED ring array for oblique illumination of the object P in two colors (red and green)?



FIGURE 3.20 Problem 3.14 - Configuration with on-axis dark field illumination.

What should be the illumination angle of each LED and what is the maximum number of LEDs in a single line ring array?

3.5. Solutions to Problems

3.1. (a) The radiation flux emitted by the filament can be calculated as follows:

$$P_{\lambda} = \varepsilon_{\lambda} e_{\rm B}(\lambda, T) \times S \frac{\omega \times \Delta \lambda}{2\pi}$$

where $S = 4.2 \times 2.3 \times 10^{-6} = 9.96 \times 10^{-6} \text{ m}^2$ is the filament irradiated surface, $\omega = 1/0.5^2 = 4$ sr is the solid angle of irradiance measurement, $\Delta \lambda = 1$ nm, and $e_{\rm B}(\lambda, T)$ is the hemispherical black body radiation at a wavelength of 0.5 µm. To calculate the last value we use tables of black body radiation (Appendix 3). Since $\lambda T = 0.5 \times 3,234 = 1,617$ we should use two lines of the table, of $\lambda T = 1,600$ and $\lambda T = 1,700$, and also take into account the factor σT^5 . Then finally we get $e_{\rm B}(\lambda, T) = 81.7 \times 5.668 \times 10^{-8} \times 3.234^5 \times 10^{15} = 163,814 \times 10^7$ W/m³ and $e_{\rm B}(\lambda, T) \times \Delta \lambda = 1,638.14$ W/m², which results in $P_{\lambda} = 0.8 \times 1,638.14 \times 9.96 \times 10^{-6} \times 4/2\pi = 8.06$ mW (at 0.5 m for 1 nm spectral bandwidth).

(b) For $\Delta U/U_0 = 0.05$ we have $\Delta T/T = 0.4(\Delta U/U_0) = 0.4 \times 0.05 = 0.02$ and therefore the new temperature is T = 3,169 K. Then we proceed using the table of Appendix 3 exactly in the same way as in (a) above and finally obtain $P_{\lambda} = 6.73$ mW.

3.2. The optical configuration addressed in this problem is presented in Fig. 3.21. Starting with the cylindrical lens we find first its optical power in the vertical plane:

$$\frac{1}{f'} = (n-1)/R_1 = 0.5163/20.65; \quad f' = 40 \text{ mm}$$



FIGURE 3.21 Problem 3.2 – A line generator with a QTH lamp.

and then the location of the line of light along the OZ axis and its height y':

$$S'_1 = \frac{40 \times 60}{(60 - 40)} = 120 \text{ mm}; \quad V = -120/60 = -2; \quad y' = 4.2 \times 2 = 8.4 \text{ mm}.$$

Obviously in the horizontal plane the lens has no optical power and the ray bundle of 1 mm width on the line has a width $\Delta x = 60/(60+120) = 0.33$ mm in the plane of the lens. Hence, all rays concentrated by the lens in the segment of area *B* of the line are those which are transferred by the lens strip $A = 20 \times 0.33 = 6.67$ mm². They constitute the solid angle $\Omega = A \cos \varphi / \rho^2$, where $\rho = 60/\cos \varphi$. For the light intensity along the created line we get

$$I(x) = \varepsilon e_{\rm B}(\lambda, T) \times \Delta \lambda \times s \times \frac{\Omega(x)}{2\pi B} \cos \varphi(x)$$

= $\varepsilon e_{\rm B}(\lambda, T) \times \Delta \lambda \times s \frac{20 \times \Delta x}{2\pi B \times 60^2} \cos^4 \varphi(x) = I_0 \cos^4 \varphi(x).$

To calculate I_0 we use the data from Problem 3.1:

$$e_{\rm B}(\lambda, T) \times \Delta \lambda = 1,628.14 \text{ W/m}^2/\text{nm}; \quad s = 9.96 \times 10^{-6} \text{ m}^2$$

and also keep in mind that $B = 1 \times y' = 8.4 \times 10^{-6} \text{ m}^2$; tan $\varphi_{\text{max}} = 150/180$; $\cos^4 \varphi_{\text{max}} = 0.348$. Therefore, the intensity along the line varies from the maximum value $I_0 = 0.439 \text{ W/m}^2/\text{nm}$ (in the center) to the minimum value $I = 0.348 \times I_0$ (at the side).

3.3. (a) As explained in Section 3.2, the two-lens configuration enables one to get a uniform illumination of the object if the first lens, L_1 , creates the image of the source S on the second lens, L_2 , and the image of lens L_1 coincides with the object plane. Such a configuration is shown in Fig. 3.22. To optimize the system with



FIGURE 3.22 Problem 3.3 - (a) Optimized illumination with two lenses and (b) configuration with the maximum useful size of the light source.

regard to illumination power incident on the object y we should pay attention to the fact that the magnifications of both lenses are reciprocal to each other:

$$S'_1 = -S_2; \quad S'_2 = -S_1; \quad V_1 = \frac{S'_1}{S_1} = \frac{-S_2}{-S'_2} = \frac{1}{V_2}$$
 (A)

and that limitation of the total length, l, yields

$$l = -2S_1 + S'_1 = -2S_1 + V_1S_1; \quad S_1 = \frac{l}{V_1 - 2}.$$
 (B)

Furthermore, the solid angle of radiation transferred from S to y is given by

$$\Omega = \left(D_{\rm eff}^{(1)}\right)^2 / \left(S_1\right)^2$$

where $D_{\text{eff}}^{(1)}$ is the effective working diameter of the first lens (see Fig. 3.22a). We should keep in mind that this value is maximized if the object y imaged back through L₂ covers all of L₁, which occurs if $V_2 = y/D$; $V_1 = D/y$. By substituting this value in Eq. (B) and then proceeding with the expression of the solid angle we obtain

$$\Omega_{\max} = \frac{D^2}{(S)_{\min}^2} = \frac{y^2}{l^2} [(V_1)_{\max}^2 (V_1 - 2)_{\max}^2] = \left[\frac{D}{l} \left(\frac{D}{y} + 2\right)\right]^2.$$

Therefore, $V_1 = -30/10 = -3$ and from Eq. (B) we get

$$S_1 = \frac{250}{-3-2} = -50 \text{ mm};$$
 $S'_1 = 150 \text{ mm};$
 $f'_1 = f'_2 = (1/150 - 1/50)^{-1} = 37.5 \text{ mm}.$



FIGURE 3.23 Problem 3.4 – (a) Imaging of the field stop D_2 and (b) the rays defining the actual size of the lens.

(b) Obviously the larger the source size *h* the greater the illumination level on the object y. An increase of *h* is accompanied by an increase of the effective working diameter of the second lens, as becomes apparent from Fig. 3.22a. Since the maximum effective size is *D*, the corresponding size of the source can be calculated as follows: $h_{\text{max}} = D/|V_1| = 30/3 = 10$ mm. This case is shown in Fig. 3.22b. It is understandable that any additional part of the source, above the size h_{max} , will be imaged outside of lens L₂ and consequently will be useless.

3.4. (a) The central issue in this problem is the location and size of the field stop D_2 . To find it we first consider the imaging branch. Using the paraxial approximation for the thin lens of f' = 40 mm we have $S'_{ob} = -2 \times S_{ob}$, and therefore Newton's formula yields $S_{ob} = -60$ mm; $S'_{ob} = 120$ mm. Then, taking into account that (i) the fiber exit is positioned in the focal plane of L_1 and therefore the rays between L_1 and L_2 are parallel; and (ii) the two lenses build the image of the bundle in the aperture D_3 with magnification $V_{21} = -12/6 = -2$, we draw the conclusion that the distance between D_3 and L_2 is equal to $f'_2 = 50$ mm and $f'_1 = 50/2 = 25$ mm. Furthermore, since the aperture D_2 is conjugate with the object plane its image created by lens L_2 should be at the same distance from L_{Ob} (and D_3) as the sensor plane y', i.e., at 120 mm from D_3 or 70 mm from L_2 (see Fig. 3.23a). Hence

$$S'_2 = \frac{70 \times 50}{120} = 29.17 \text{ mm}; \quad V_2 = \frac{29.17}{70} = 0.417;$$

 $D_2 = D'_2 V_2 = 4 \times \sqrt{2} \times 0.417 = 2.35 \text{ mm}.$

(b) The uppermost tilted beam passing through D_2 is determined by the angle β corresponding to the side point of the fiber bundle. A geometrical consideration of Fig. 3.23b gives $D_{L1} = 2h = 2(D_2/2 + 29.17 \tan \beta) = 9.35 \text{ mm} = D_{L2}$. What remains to check is that the whole segment 2h is illuminated by the fiber.



FIGURE 3.24 Problem 3.5 – Illumination in a luminescent microscope.

Indeed, the lowest point M illuminated by the fiber has the vertical coordinate $y_{\rm M} = 3 + 25 \tan(\arcsin 0.25) = 9.45$ mm, which is greater than *h*.

3.5. (a) We choose the configuration shown in Fig. 3.24, and perform calculations in the paraxial approximation. Beginning with the magnification of the second lens, we get

$$V_2 = -\frac{\text{FOV}}{D_1} = -0.18.$$

Then the geometry of the principal ray yields:

$$\tan \alpha = \tan[\arcsin(\text{NA})] = -\frac{a}{S_1} \times \frac{S_1'}{S_2'} = \frac{a}{S_1 V_2}$$

which results in

$$S_1 = \frac{a}{\tan \alpha \times V_2} = \frac{0.5}{0.185 \times (-0.18)} = -15.0 \text{ mm}$$
$$S'_1 = \frac{S_1 f'_1}{f'_1 + S_1} = 97.5 \text{ mm} = -S_2.$$

Therefore, the focal length of the second lens is

$$f_2' = \frac{S_2 V_2}{1 - V_2} = 14.9 \text{ mm}$$

and $S'_2 = S_2V_2 = 17.55$ mm; $D_2 = 2aV_1 = 1 \times 97.5/15 = 6.5$ mm. Finally, taking into account that $S_3 = -S'_2$ and that L_3 images lens L_2 and the source S into the entrance pupil P which is 18 mm from the object, we obtain the focal length of L_3 :

$$f'_3 = \frac{18 \times 17.55}{18 + 17.55} = 8.9 \text{ mm.}$$

(b) The fraction of the source radiation which is transferred by the illumination system is dictated by the solid angle $\omega = (D_1/S_1)^2 = 0.444$ sr. Since irradiance E_{λ} is usually measured at a distance of 0.5 m and averaged over 1 m² area, e.g., in the solid angle $\Omega = 1/0.25 = 4$ sr, we can calculate the optical power P_{λ} incident on the full FOV in the spectral range of 1 nm as follows:

$$P_{\lambda} = E_{\lambda} \frac{\omega}{\Omega} = 28 \times 0.444/4 = 3.1 \text{ mW/nm}.$$

The cell of 6 μ m in size obtains the portion (6/1, 800)² = 11.1 × 10⁻⁶ of that power, i.e., 3.45 × 10⁻⁸ W. Since each photon of radiation of 240 nm transfers the energy

$$E_{\rm Ph} = \frac{hc}{\lambda} = \frac{6.625 \times 10^{-34} \times 3 \times 10^8}{0.24 \times 10^{-6}} = 0.83 \times 10^{-18} \,\mathrm{J}$$

one can calculate the number of photons per second incident on the cell: $N = P_{\lambda}/E_{\text{Ph}} = 4.16 \times 10^{10}$ photons/s.

3.6. Using Eq. (3.4) one can obtain

$$w_0 = \sqrt{\frac{0.6328 \times 10^{-3} \times 300}{2\pi}} = 0.174 \text{ mm}$$

at the waist of the beam in the middle of the cavity. Hence, the full diameter at the exit of the laser is $D = 2w = 2\sqrt{2}w_0 = 0.49$ mm. The angle of divergence is calculated from Eq. (3.3) as follows:

$$2\theta = \frac{2\lambda}{\pi w_0} = 2.32 \text{ mrad.}$$

3.7. First we consider the general expression for radiation power incident on the area of the detector positioned at a coordinate *z*. Denoting the radius of the detector as r_d and $P_0 = 2$ mW we have from Eq. (3.5)

$$P_{z} = \pi I_{0} \int_{0}^{r_{d}} 2r \exp\left(-\frac{2r^{2}}{w_{z}^{2}}\right) dr = 2 \frac{P_{0}}{w_{z}^{2}} \int_{0}^{r_{d}^{2}} \exp\left(-\frac{2x}{w_{z}^{2}}\right) dx$$
$$= P_{0} \left[1 - \exp\left(-\frac{2r_{d}^{2}}{w_{z}^{2}}\right)\right].$$

Therefore

$$\frac{2r_{\rm d}^2}{w_z^2} = -\ln\left(1 - \frac{P_z}{P_0}\right).$$
 (A)

Taking into account that

$$w_z^2 = w_0^2 \left[1 + \left(\frac{\theta z}{w_0}\right)^2 \right] = w_0^2 + \theta^2 z^2$$
 (B)

and calculating from Eqs. (3.3) and (3.4)

$$w_0 = \sqrt{\frac{0.63 \times 10^{-3} \times 300}{2\pi}} = 0.173 \text{ mm}; \quad \theta = \frac{0.63 \times 10^{-3}}{\pi 0.173} = 1.16 \times 10^{-3},$$

we get $w_z^2 = 0.03 + 1.346 \times 10^{-6} z^2$ and after substituting in Eq. (A):

$$0.03 + 1.346 \times 10^{-6} z^2 = -\frac{2r_d^2}{\ln(1 - P_z/P_0)}.$$
 (C)

The maximum distance z_{max} is related to $P_z = \text{NEP} = 10^{-7}$ which yields $(2r_d^2 = 1.125 \text{ mm}^2)$: $z_{\text{max}}^2 = 1.68 \times 10^{10} \text{ mm}^2$; $S_{\text{max}} = z_{\text{max}} - L/2 = 129.85 \text{ m}$. The minimum distance is related to the value $P_z = \text{NEP} \times 10^4 = 10^{-3}$ and again by substituting in Eq. (C) we get $z_{\text{min}}^2 = 1.184 \times 10^6 \text{ mm}^2$; $S_{\text{min}} = z_{\text{min}} - L/2 = 938 \text{ mm}$.

3.8. (a) Let us find first the beam waist radius, w_0 , and the Rayleigh length, z_R , using Eq. (3.4):

$$w_0 = \sqrt{\frac{\lambda L}{2\pi}} = \sqrt{\frac{0.63 \times 0.3}{2\pi}} = 0.174 \text{ mm}; \quad z_{\rm R} = \frac{\pi w_0^2}{\lambda} = 150 \text{ mm}.$$

Taking into account the Gaussian profile of the laser beam we use Eq. (3.7) to find the distance S' between the lens and the beam waist. In the first case S = -(150 + 5) = -155 mm:

$$\frac{1}{S'} = \frac{1}{f'} + \frac{1}{S + z_{\rm R}^2/(S + f')} = \frac{1}{80} - \frac{1}{155 + 150^2/(-155 + 80)}; \quad S' = 97 \text{ mm}.$$

In the second case S = -(150 + 100) = -250 mm; and the same calculation of S' as above gives S' = 101 mm. Hence, the distance that the waist moves is 4 mm greater than that of the lens moving.

(b) To consider the general situation we introduce the function

$$q(x) = x + \frac{z_{\rm R}^2}{x + f'}$$

and find its maximum, as usual, by analyzing the derivative

$$\frac{\mathrm{d}q}{\mathrm{d}x} = 1 - \frac{z_{\mathrm{R}}^2}{(x+f')^2}$$

which has a zero value at $x + f' = -z_R$. Therefore $S = x = -(z_R + f') = -230$ mm corresponds to the maximum achievable distance S':

$$S'_{\text{max}} = \left(\frac{1}{80} - \frac{1}{230} \times \frac{1}{1 + 150^2/(230 \times 150)}\right)^{-1} = 101.32 \text{ mm}.$$

3.9. We begin with a calculation of the laser beam parameters and use Eqs. (3.3) and (3.4):

$$w_0 = \sqrt{\frac{0.63 \times 10^{-3} \times 500}{2\pi}} = 0.224 \text{ mm}; \ 2\theta = \frac{0.63 \times 10^{-3}}{\pi 0.224} = 1.79 \times 10^{-3}.$$

The beam diameter at the exit of the laser (and also at the entrance of the beam expander) is $2w = 2w_0\sqrt{2} = 0.632$ mm. Now, by substituting in Eq. (3.9) the values $W = 10^{-1}$ and l = 3,000 one can find the beam diameter at the side of the robot detector:

$$D' = 0.632 \times 10 + \frac{3,000 \times 1.79 \times 10^{-3}}{10} = 6.87 \text{ mm.}$$

Since this value is smaller than the robot lens diameter, the optical power registered by the detector will be determined by the actual size of the lens:

$$P = P_0\{1 - \exp[-2 \times (10/6.87)^2]\} = 5 \times 0.986 = 4.92 \text{ mW}.$$

3.10. (a) The divergence angle at the laser exit, before the beam expander, can be found from Eq. (3.3):

$$2\theta = 2\frac{\lambda}{\pi w_0} = 2\frac{0.514 \times 10^{-3}}{\pi \times 0.7} = 0.468$$
 mrad.

Hence, after the beam expander of angular minification $W = (100/8)^{-1} = 12.5^{-1}$ we get $2\theta' = 37.4 \times 10^{-6}$.

(b) If the spatial filter is positioned just after the laser the pinhole size is calculated from Eq. (3.10) (we choose the factor c = 4/3 which is commonly accepted):

$$d_{\rm SF}^{(1)} = \frac{4}{3} 2\theta f' = 12.5 \,\,\mu{\rm m}.$$

In the second case the beam diameter just after the beam expander is (assuming lens L₂is large enough) $D_2 = 2w_0\sqrt{2}/W = 2 \times 0.7 \times \sqrt{2} \times 12.5 = 24.7$ mm, which is larger than the spatial filter lens of 10 mm. Therefore, the pinhole size should be calculated by considering diffraction of the Gaussian beam truncated by lens L₃, namely:

$$d_{\rm SF}^{(2)} = \frac{4}{3} \times \frac{1.22\lambda}{D_3} f_3' = \frac{4}{3} \times \frac{1.22 \times 0.568 \times 20}{10} = 1.7 \,\mu{\rm m}.$$



FIGURE 3.25 Problem 3.11 – Influence of a second lens on beam divergence.

(c) The optical power P_3 transferred by the spatial filter in the second case is calculated by integrating Eq. (3.5) from 0 to $D_3/2 = 5$ mm. This yields

$$P_3 = P_0\{1 - \exp[-2(D_3/D_2)^2]\} = P_0 \times 0.28.$$

Thus, only 28% of the total power of the laser will pass the spatial filter.

3.11. The divergence of the exit beam from the transmitter with a single lens L_1 is 1 mrad which gives, after traveling a distance of 200 m, a beam size of about 200 mm, i.e., much greater than the receiver detector (see Fig. 3.25). If the second lens, L_2 , is added to the transmitter the angle of the side ray has to be limited by the detector size as follows:

$$D = l = D_0 - 2\alpha' S;$$
 $\alpha' = -\frac{l - D_0}{2S} = -\frac{50 - 3}{400 \times 10^3} = -1.175 \times 10^{-4}.$

Remembering the relation between the angles before and after the simple lens (see Section 1.1) and keeping in mind the sign convention, we get

$$\Phi_2 = \frac{\alpha' - \alpha}{h} = \frac{(-1.175 + 5) \times 10^{-4}}{1.5} = 2.55 \times 10^{-4}; \quad f_2' = 3,922 \text{ mm}.$$

3.12. (a) We should first locate the laser diode with regard to lens L₁. As an initial approximation one can start with the OYZ plane (see Fig. 3.26a) and use the paraxial formula for the lens which is a complete circle of 7 mm in diameter in this plane, and therefore f' = 5.136 mm (assuming n = 1.5 for the glass). Each principal plane, H and H', is 3.5 mm inside the lens. This gives $l_p = 5.136-3.5 = 1.636$ mm, if the laser output is just in the paraxial focus. We choose $l_p = 1.5$ mm to start a trial and error procedure aimed at obtaining the minimum achievable radius of the beam, w_y , in the plane K. Exploiting the precise ray tracing formula for a cylindrical lens as in Problem 2.6, one can use for $u = 3.5^{\circ}$ the results obtained in the solution of that problem: $y' = w_y = 4.57$ mm. Then we proceed in the same way, but trying the value l = 2.0 mm. The ray tracing procedure gives $w_y = -4.2$ mm. Then, for l between the two previous trials, e.g., l = 1.75 mm,



FIGURE 3.26 Problem 3.12 – Ray tracing of the line generator in (a) the OYZ plane and (b) the OXZ plane.

the ray tracing yields $w_y = 0.11$ mm and we stop the trial and error procedure at this value which is very small.

Now we consider the beam in the second plane, OXZ (Fig. 3.26b). Since the cylindrical lens here acts like a slab of glass with no optical power one can find for the angle $u = 15^{\circ}$: $w_x = 1,500 \times \tan 15^{\circ} = 401.9$ mm. The intensity distribution in plane K is described by Eq. (3.2) where two terms, that of OX and that of OY, are separated (due to ellipticity of the laser diode beam):

$$I_{\rm L}(x,y) = I_0 \exp\left[-\left(\frac{2x^2}{w_x^2} + \frac{2y_2}{w_y^2}\right)\right]$$
(A)

where

$$I_0 = \frac{2P}{\pi w_x w_y} = \frac{2 \times 10}{\pi \times 0.11 \times 401.9} = 0.144 \text{ mW/mm}^2.$$

(b) We define the contrast in plane K, $C_{\rm K}$, and in the plane of the CCD, $C_{\rm CCD}$, as follows:

$$C_{\rm K} = \frac{I_{\rm L} + I_{\rm Sun}}{I_{\rm Sun}}; \quad C_{\rm CCD} = C_{\rm K} \tau_{\rm F}$$
(B)

where $\tau_{\rm F}$ is the transmittance of the interference filter (see Eq. (5.36) of Chapter 5). Assuming that the filter has maximum transmittance at 680 nm (wavelength of the laser) for the rays which are normal to its surface, we calculate the shift in the maximum wavelength for the side direction ($u' = 15^{\circ}$): $\lambda' = \lambda \cos 15^{\circ} = 670$ nm. This means that the transmittance curve of the filter is moved left along the wavelength scale, as demonstrated in Fig. 3.27. Since the laser wavelength remains the same for the center point as well as for the side, it is apparent that one has to find the value $\tau_{\rm F}$ for the point located 10 nm to the side of the maximum.



FIGURE 3.27 Problem 3.12 – Transmittance curve at normal incidence and for a tilted beam.

From Eq. (5.36) we have

$$\frac{\Phi}{2} = \frac{\pi}{\lambda} 2nt \cos r = \frac{\pi}{\lambda} m\lambda' = \pi m \left(1 - \frac{\Delta\lambda}{\lambda}\right);$$
$$\sin^2\left(\frac{\Phi}{2}\right) = \left(\pi m \frac{\Delta\lambda}{\lambda}\right)^2 = \left(\pi \frac{\lambda}{\delta\lambda \times N_e} \times \frac{\Delta\lambda}{\lambda}\right)^2$$

and therefore

$$\tau_{\rm F} = \tau_{\rm max} \frac{1}{1 + \frac{4R}{(1-R)^2} \left(\frac{\pi}{N_{\rm e}}\right)^2 \left(\frac{\Delta\lambda}{\delta\lambda}\right)^2} = \tau_{\rm max} \frac{1}{1 + 4 \times \left(\frac{\Delta\lambda}{\delta\lambda}\right)^2}$$
$$= \tau_{\rm max} \frac{1}{1 + 4 \left(\frac{10}{10}\right)^2} = \frac{\tau_{\rm max}}{5}$$

(we also used in the last transformation Eqs. (5.38) and (5.39) from Chapter 5).

Now we have to compute the sun radiation coming to plane K at the bandwidth of 10 nm. We proceed as in Problem 3.1, using the table of black body radiation at T = 6,000 K, and finding for $\lambda T = 0.68 \times 6,000 = 4,080$ that $e_{B\lambda}/\sigma T^4 =$ $17.689 \times 10^{-5} \times T = 1,061.34/\mu m$. Keeping in mind that at the earth's surface only a portion $(1,350/\sigma T^4)$ of all radiation emitted by the sun is received, we find for the sun radiation at sea level for a bandwidth of 1 μ m: $\tilde{e}_{B\lambda} = 1,350 \times$ 1,061.34 = 1.433 mW/mm²/ μ m; and finally for the bandwidth of 10 nm: $\tilde{e}_{B\lambda} \times$ $\Delta\lambda = 1.433 \times 10^{-2}$ mW/mm² = I_{Sun} . By substituting this value in Eq. (B) we get the contrast at the CCD plane at the center point:

$$C_{\text{CCD}}(0) = C_{\text{K}} = \frac{14.4 + 1.433}{1.433} = 11.05$$

and the contrast at the CCD plane at the side point (obviously the amount of sun radiation transmitted by the filter remains the same for normal incidence and for the tilted rays):

$$C_{\text{CCD}}(w_x; 0) = \frac{I_0 \exp(-2) \times \tau_{\text{max}}/5 + I_{\text{Sun}} \times \tau_{\text{max}}}{I_{\text{Sun}} \times \tau_{\text{max}}} = \frac{1.95/5 + 1.433}{1.433} = 1.27.$$

Therefore, the contrast of the image at the side point is reduced drastically compared to the center.

3.13. Two prisms, identical to each other with regard to their shape, size, and deviation angle α , are arranged as depicted in Fig. 3.28 in the path of a collimated elliptical laser beam. Suppose AB is the minimum principal diameter of the beam before the prism pair and MN is the same segment after the prisms. To simplify the consideration we choose the spatial position of the first prism in such a manner that the rays refracted at the first surface propagate in the direction normal to the second surface and therefore both incident angle and refraction angle here are equal to zero. The same is true for the last surface of the second prism.

Our goal is to find the incident and the refraction angles at all surfaces of the prism pair $(i_1, r_1, i_2, r_2, i_3, r_3, i_4, r_4)$ and to calculate the ratio R = MN/AB which evidently can be expressed as follows:

$$R = \frac{\mathrm{MN}}{\mathrm{AB}} = \frac{\mathrm{MN}}{\mathrm{KL}} \times \frac{\mathrm{CD}}{\mathrm{AB}} = \frac{\cos r_1 \times \cos r_3}{\cos i_1 \times \cos i_3} = \left(\frac{\cos r_1}{\cos i_1}\right)^2 = R_1^2.$$
(A)

We take into account here that $\cos i_2 = \cos r_2 = \cos i_4 = \cos r_4 = 1$ and therefore

$$r_1 = \alpha. \tag{B}$$

Thus,

$$R_1 = \sqrt{R} = \frac{\cos \alpha}{\cos i_1} \tag{C}$$



FIGURE 3.28 Problem 3.13 – Anamorphic prism pair.

where i_1 is defined by Snell's law of refraction, i.e., $\sin i_1 = n \sin \alpha$. Therefore, Eq. (C) is the non-linear equation with regard to α for any given value *R*. We choose n = 1.5 and use the trial and error approach in order to solve the equation $R_1(\alpha) = \sqrt{2.0}$. Starting with $\alpha = 30^\circ$ (which yields $R_1 = 1.31$) we proceed further until the final value $\alpha = 32.3^\circ$ obeys the equation with proper accuracy. The corresponding value of the first incident angle is $i_1 = 53.3^\circ$ and the same is true for i_3 which defines the spatial position of both prisms with regard to the horizontal axis.

3.14. Starting with the imaging optics, we use the paraxial approximation formula for lens L_1 :

$$\frac{1 - V_1}{S_1 V_1} = \frac{1}{f'_1}; \quad S_1 = -25 \frac{1.25}{0.25} = -125 \text{ mm}; \quad S'_1 = 31.2 \text{ mm}$$

and the field of view (FOV) in the plane P is $25.5 \text{ mm} \times 19.2 \text{ mm}$ with the diagonal (maximum size) of 32 mm. Therefore, the illumination branch should provide light to any point inside the circle of 32 mm.

The dark field effect will be achieved if radiation coming to the plane P is reflected specularly and collected by lens L_1 on the CCD. Correct illumination requires that the image of the light source (LED) will be obtained in the plane of L_1 . Taking into account that the maximum illumination angle is dictated by the condenser diameter (45 mm) and that the focal length of L_2 is also 45 mm (f # = 1.0), we consider the ray trajectories in the system as shown in Fig. 3.29a where the unfolded version of the illumination branch is shown by dotted lines.



FIGURE 3.29 Problem 3.14 – (a) Layout of imaging and illumination optics and (b) unfolded diagram of the illumination branch.

Then, keeping in mind the required working distance, we have

$$S'_2 = 45 + 16 + 125 = 186 \text{ mm};$$
 $S_2 = -\frac{45 \times 186}{141} = -59.4 \text{ mm};$
 $V_2 = -3.13;$ $a = 186 - 125 - 45/2 = 38.5 \text{ mm};$ $y'_2 = 3 \times 3.13 = 9.4 \text{ mm};$

where y'_2 is the active size of lens L₁. Thus, the location of all elements is determined and to find the size of the beam splitter, AB, we proceed further with Fig. 3.29b as follows: AO + OB = $(22.5 + z)\sqrt{2}$; where for z we have

$$\frac{z-4.7}{125+38.5-z} = \frac{22.5-4.7}{186}; \quad z = 18.57 \text{ mm and } AB = 60 \text{ mm}.$$

Finally, we have to estimate vignetting. From Fig. 3.29b it becomes evident that the rays from all points of the source are transferred through each point of the segment DD_1 (no vignetting) and up to 50% of the rays are transferred through each point of the segments CD and C_1D_1 (vignetting of 50% or less). This means that at full FOV (segment $CC_1 = 32$ mm) vignetting does not exceed 50%.

3.15. To ensure that no additional vignetting will be added to the system described in Problem 3.14, we introduce a stop ST at a distance of 16 mm from the object and find the size of ST using the unfolded version of the illumination branch (see Fig. 3.29b):

$$D_{\rm ST} = 32 \times (125 + 16)/125 = 36$$
 mm.



FIGURE 3.30 Problem 3.15 – (a) Layout of LEDs in oblique illumination and (b) arrangement of the ring array.

Furthermore, we choose LEDs with embodied lenses, the outer diameter of each LED being 5 mm and operating at a wavelength of 535 nm for green illumination and 630 nm for red. The viewing angle will be found later.

All LEDs should be located on a circle of diameter 46 mm at least and each one should be tilted towards the object P by an angle φ , as depicted in Fig. 3.30a. Assuming that the source from the left side illuminates the left part of the field of view (FOV) and the corresponding LED from the right side illuminates the right part of the FOV, we get from the geometry shown in the figure: (i) the viewing angle of a single LED

$$\beta = \psi_2 - \psi_1 = \arctan\left(\frac{46/2}{16}\right) - \arctan\left(\frac{46/2 - 16}{16}\right) = 32^\circ;$$

and (ii) the angle of tilting

$$\varphi = \arctan\left(\frac{46/2 - 8}{16}\right) = 41.3^{\circ}.$$

Also, choosing the spacing between two adjacent LEDs to be 4 mm (LED pitch of 9 mm along the ring circle), we can calculate the number of LEDs in the ring: $N = int (\pi 46/9) = 16$.

Detectors of Light

4.1. Classification of Radiation Detectors and Performance Characteristics

As was mentioned earlier, radiation detectors act as transformers converting energy of incident photons into energy of electric carriers or, simply, into electrical signals (current or voltage). There exists a great variety of radiation detectors different in their physical basis, hardware realization, and performance characteristics. Each electro-optical system requires detectors which optimally suit the specific application.

Commonly used radiation detectors can be classified as follows:

- (a) physical process of signal generation
 - electro-optical detectors single electro-optical cells photomultipliers
 - semiconductor detectors photoconductive detectors (photoresistors) photovoltaic detectors (photodiodes)
 - thermal detectors (bolometers)
- (b) a number of detectors in a single housing
 - single detectors
 - detector arrays line detectors (one-dimensional arrays) area sensors (two-dimensional arrays).

It should be mentioned that there exist optical sensors of different kinds and configurations not presented in the above classification – we address here only the most popular radiation detectors.

There are a number of parameters characterizing each and any detector. We describe here the most important of them.

Responsivity, $R_{\lambda} = di_d/dE$, is measured in A/W, i.e., current of the detector, i_d (in amperes) per watt of incident optical power. Obviously responsivity is a spectrally dependent value since a detector is sensitive (meaning it is capable of absorbing photons and generating a corresponding amount of electrons) to radiation within some finite interval of wavelengths. The wavelength interval where R_{λ} is still noticeable is called the working spectral range of the detector.

Quantum efficiency, $\eta = N_e/N_{\rm ph}$, is defined as the ratio of the number of generated electrons, N_e , to the number of incident photons, $N_{\rm ph}$. Usually $\eta < 1$ and of course it also depends on wavelength. Quantum efficiency and responsivity are evidently related to each other. Indeed, if we take into account that $N_e = i_d/e = R_\lambda E_\lambda d\lambda/e$ (*e* is the electrical charge of a single electron) and $N_{\rm ph} = E_\lambda d\lambda/(hc/\lambda)$, then, by inserting $hc/e = 1.240 \times 10^{-6}$ Wm/A, one can obtain from the definition of quantum efficiency

$$\eta_{\lambda} = \frac{1.24}{\lambda} R_{\lambda} \tag{4.1}$$

where the wavelength λ should be in μ m.

Noise equivalent power (NEP). In order to define NEP we have to consider the noise which always occurs in the detector itself and in the detector's electric circuitry. The reasons for and nature of noise will be considered later. What is important here is the fact that any useful signal generated in the detector is always accompanied by randomly varying noise. Signal-to-noise ratio (SNR) determines how strong is the signal compared to noise. The SNR should be as high as possible. For weak signals the SNR is of the order of 3:1 or 2:1 and as a limiting situation it is chosen that SNR = 1:1. The incoming power corresponding to such a limiting case is called the NEP. Hence, NEP is the minimum radiation still capable of detection. It is obviously measured in watts.

Detectivity, D, is defined just as the reciprocal of NEP:

$$D = 1/\text{NEP}$$
 (measured in W⁻¹) (4.2)

or, keeping in mind that NEP is related to noise, i_n , which is a random function of time and therefore is characterized by its mean value, $\sqrt{i_n^2}$, we get

$$D = \frac{R_{\lambda}}{\sqrt{\bar{i}_{n}^{2}}}.$$
(4.3)



FIGURE 4.1 Detector signal resulting from an impulse of incident radiation.

Specific detectivity, D^* , is similar to detectivity, but, in addition to NEP, it also takes into consideration the detector active area, A, and the bandwidth, Δf , i.e., the range of frequencies to which the detector and its electric circuitry are capable of responding:

$$D^* = \frac{\sqrt{A \times \Delta f}}{\text{NEP}}, \text{cm}^{1/2} \text{Hz}^{1/2} \text{W}^{-1}.$$
(4.4)

Obviously *A* is measured in cm² and the bandwidth in Hz. The value D^* (also called "D-star") is useful when a comparison of detectors of different sizes operated at different frequencies is required. Values of D^* of 10^{10} to 10^{13} are usual for this parameter.

Time response. Detection of radiation that varies in time requires that the detector circuitry will be fast enough, or, in other words, will be of suitable frequency bandwidth. Several parameters are related to the time response of the detector, as is evident from Fig. 4.1. Radiation is incident on the detector during a very short time interval, δ . Due to electrical inertial processes in the detector and in the elements of its circuitry the generated current has a finite speed of growth characterized by the rise time, τ_{Rise} , as well as a finite speed of decay characterized by the fall time, τ_{Fall} . Both values, τ_{Rise} and τ_{Fall} , are determined with regard to half the maximum of the generated signal, as is shown in Fig. 4.1, and the maximum working frequency can be calculated as

$$f_{\rm max} = \frac{1}{2(\tau_{\rm Rise} + \tau_{\rm Fall})}.$$

Dark current, $i_{d.c.}$. Even if no radiation is incident on a detector connected to some electrical circuitry, the output of the circuitry is not equal to zero but has some finite value called the dark current. This results from physical processes inside the detector and elements of the circuitry. Since dark current is a fluctuating process it can be characterized by a DC component, but also by mean fluctuation value, $\overline{i}_{d.c.}^2$.
Dynamic range (DR) is defined as the ratio between the maximum and the minimum detectable radiation:

$$DR = \frac{E_{\max}}{E_{\min}} = \frac{i_{d \max}}{i_{d \min}}$$
(4.5)

(it is assumed that the detector is operated in its linear range when i_d is proportional to the incident radiation power). The maximum value in Eq. (4.5) is dictated by the saturation of the detector, meaning that at some level of radiation all available mobile carriers of charge (electrons) are already generated and no additional electrons can be created if additional photons arrive at the detector. The minimum value in Eq. (4.5) is governed by the detector noise and usually it is equal to NEP. In the right-hand term of Eq. (4.5) the dark current is often exploited as the minimum detector signal.

Problems

4.1. Find the responsivity of a photodetector made of GaAs which has maximum sensitivity at a wavelength of 0.83 μ m with a quantum efficiency of 10%.

4.2. Which one of two detectors available for a laboratory set-up will generate greater current for the same illumination conditions and measured frequencies?

Detector 1: $\eta = 0.3$;	$NEP = 2 \times 10^{-14} \text{ W/Hz}^{1/2}$
Detector 2: $\eta = 0.5$;	$NEP = 4 \times 10^{-15} \text{ W/Hz}^{1/2}$

Both detectors are equivalent with regard to figure of merit (specific detectivity).

4.3. Calculate the detectivity and specific detectivity of a silicon photodetector of 3 mm² area for maximum response wavelength ($\lambda = 0.86 \,\mu\text{m}$; $\eta_{\lambda} = 0.83$), if the measured noise is 2×10^6 e/ms at an operation bandwidth of 100 kHz.

4.4. Investigating the dynamic features of a detector it is found that the rise time is 2.5 times less than the fall time, both being measured at a response to very short radiation pulses (of duration δ). The time constant of the detector output circuitry is estimated as $t_0 = 0.5$ ns. Find the operation bandwidth of the system.

4.2. Noise Consideration

There are a number of reasons for an output detector signal not being constant, but varying randomly (fluctuating) with time. These random fluctuations, defined as

the detector noise, are superimposed on the useful signal and obviously influence the performance of the device, especially its ability to detect weak radiation or to differentiate between two very close levels of radiation intensity.

From the mathematical point of view noise is considered as a random function and should be characterized by statistical parameters, as is usual in such a case (mean value, mean square value, standard deviation, correlation functions, etc.). We will address here three main kinds of noise: shot noise, thermal (Johnson) noise, and read-out noise.

Shot Noise

This results from the discrete nature of carriers of radiation energy (photons) and carriers of electric charge (electrons or holes) generated in a detector.

Since photons are created by a light source in a random manner, the number of photons, N_{ph} , arriving at the detector during the time interval *T* is not constant, but varies from one time interval to another. The probability of finding *N* photons coming to the detector obeys the Poisson distribution law:

$$P(N_{\rm ph}) = \frac{(\overline{N}_{\rm ph})^{N_{\rm ph}}}{N_{\rm ph}!} (\mathrm{e}^{-\overline{N}_{\rm ph}})$$
(4.6)

where \overline{N}_{ph} is the mean value (averaged over time *T*) and it can be shown that for the mean fluctuation of the process the following relation is valid:

$$\sigma_{\rm ph}^2 = \overline{(N_{\rm ph} - \overline{N}_{\rm ph})^2} = \overline{N}_{\rm ph}.$$
(4.7)

It is important to notice that the Poisson distribution (Eq. (4.6)) appears as a result of considering a situation where the probability of a single event $p = n_1 dt$ and, therefore, if dt is approaching zero (infinitesimally small time interval) the number of independent statistical tests, N, in time T is increasing to infinity. The value n_1 (the number of photons in a time unit) remains constant, as does the mean number of photons in time $T: \overline{N_{\text{ph}}(T)} = n_1 T$.

Since the number of photons impinging on the detector, N_{ph} , limits the number of statistical tests in the consideration of generated charge carriers (electrons or holes), the statistics in this case obeys the Bernoulli distribution:

$$P(N_{\rm e}) = \frac{N_{\rm ph}!}{N_{\rm e}! \times (N_{\rm ph} - N_{\rm e})!} p^{N_{\rm e}} (1 - p)^{N_{\rm ph} - N_{\rm e}}$$
(4.8)

with standard deviation (mean fluctuations) expressed as

$$\sigma_{\rm e}^2 = (\overline{N_{\rm e} - \overline{N}_{\rm e}})^2 = \overline{N}_{\rm e}(1 - p). \tag{4.9}$$

The probability *p* in this case is just the quantum efficiency of the detector $(p = \eta)$ and also $\overline{N}_{e} = \eta \overline{N}_{ph}$. Hence, for the total fluctuation of carriers caused by both processes, variation of incoming photons and variation of generated electrons for each given number of photons, we have

$$\overline{(N_{\rm e} - \overline{N}_{\rm e})_{\rm total}^2} = \overline{(N_{\rm ph} - \overline{N}_{\rm ph})^2} \times \eta^2 + \overline{(N_{\rm e} - \overline{N}_{\rm e})^2} = \overline{N}_{\rm ph} \eta^2 + \overline{N}_{\rm ph} \eta (1 - \eta) = \overline{N}_{\rm e}.$$
(4.10)

Keeping in mind that the fluctuation of a number of carriers is related directly to the detector current noise as

$$\overline{i_{\text{Sn}}^2} = \overline{(N_{\text{e}} - \overline{N}_{\text{e}})_{\text{total}}^2} \times \frac{e^2}{\tau^2} = (\overline{N}_{\text{e}}e/\tau)\frac{e}{\tau} = i_{\text{d}}e/\tau$$

and converting the time interval τ to the corresponding frequency bandwidth $\Delta f = 1/2\tau$ one obtains the well-known formula for shot noise:

$$i_{\rm Sn} = \sqrt{i_{\rm Sn}^2} = \sqrt{2i_{\rm d}e \times \Delta f} \tag{4.11}$$

where *e* is the electron charge and the detector current i_d should include the signal and also the dark current.

Thermal (Johnson) Noise

This is caused primarily by temperature fluctuations in the electrical resistance of the detector and/or the load resistor of the detector circuitry. Denoting the relevant resistance as R_L the corresponding expression for thermal noise is

$$i_{\rm Tn} = \sqrt{i_{\rm Tn}^2} = \sqrt{\frac{4kT}{R_{\rm L}}}\Delta f \tag{4.12}$$

or in terms of voltage measured on the load resistor (see Fig. 4.2)

$$\sqrt{\overline{V_{\rm Tn}^2}} = \sqrt{4kTR_1 \times \Delta f}.$$
(4.13)



FIGURE 4.2 Schematic of a detector load resistor.

If both shot noise and Johnson noise are present simultaneously in the system the optimal load resistance is usually defined as the one which causes both noise components to be equal.

Read-out Noise

This kind of noise occurs in detector arrays, like CCD or CMOS sensors, where the signals of separate elements are transferred to a single output from which they are read out sequentially, one by one. There are several origins of noise in such a readout procedure. First of all, it results from the dark current caused by thermally generated electrons in each element of the array. This process is exponentially dependent on temperature

$$i_{\rm d.c.} = \operatorname{const} \times \exp\left(-\frac{U_{\rm G}e}{2kT}\right)$$
 (4.14)

where $U_{\rm G}$ is the gap in the energy diagram of the pixel material (usually silicon semiconductor) and *e* is the electron charge. There is a special read-out protocol allowing one to avoid the influence of the (averaged) dark current on the signal pattern registered by the array, but differences between elements, which inevitably exist in any array of detectors, result in so-called fixed pattern noise and dark current non-uniformity. Both have an impact on the read-out noise, as well as several other sources acting in the electrical circuitry of the device.

It is the read-out noise r.m.s. value that limits the low end of the registered radiation power and it is this value that is frequently used in the calculation of the dynamic range of the detector array. Read-out noise r.m.s. is measured usually as the number of electrons per read-out sequence. Obviously effective cooling of a device is capable of reducing drastically the dark current and the read-out noise.

Problems

4.5. (a) Find the minimum flux of photons which can be detected by a sensor of NEP = 10^{-9} W in visible light operated at a frequency of 50 MHz. (b) Might a fluctuation of 20% be reasonably observed in this flux?

4.6. Assuming the gap of silicon to be $U_G = 1.1$ V, calculate the possible reduction in dark current of a CCD if it is cooled from room temperature (25°C) to 0°C.

4.7. A CCD array of $7 \times 7 \mu m$ pixel size is operated at a video rate (30 frames per second). It has saturation exposure $E_{\text{sat}} = 0.2 \,\mu \text{J/cm}^2$, quantum efficiency $\eta = 0.2$, and read-out noise of 50 electrons. What is the dynamic range of the CCD?

4.3. Single Electro-optical Detectors (Photocells, Photomultipliers, Semiconductor Detectors, Bolometers)

Photoelectric Cells

A photoelectric cell consists of two electrodes, cathode and anode, placed in an evacuated vessel (tube) transparent for incoming radiation (see Fig. 4.3). An external voltage source provides an appropriate voltage drop, V, between the electrodes. Operation of the cell is based on the photoelectric effect obeying Einstein's equation:

$$E_{\rm ph} = h\nu = W_{\rm es} + \frac{mV^2}{2}$$
 (4.15)

which states that the incident photon energy is equal to the sum of the photoelectric work function (work of escape), W_{es} , necessary for an electron to escape from the photocathode and the kinetic energy of the electron just after leaving the electrode. The work function depends on the properties of the photocathode material (for example, for Cs it is 1.8 eV whereas for Ge it is 4.5 eV). It is understandable that W_{es} causes a limitation on the wavelengths at which the photoelectric effect (and therefore the electron current in the cell) can be obtained, namely: $(E_{ph})_{min} = W_{es}$. As a result, a maximum wavelength (sometimes called the threshold wavelength) exists that is still capable of releasing the electron:

$$\lambda_{\max} = \frac{hc}{W_{\text{es}}} = \frac{1.240}{W_{\text{es}}} \tag{4.16}$$

 $(W_{es} \text{ in eV} \text{ and wavelength in } \mu \text{m})$. This means that for any wavelength larger than the value of Eq. (4.16) there is no way to get a photocurrent, neither by increasing the voltage drop nor by concentrating more photons on the photocathode.



FIGURE 4.3 Schematic of a photocell.



FIGURE 4.4 Typical graph of photocurrent vs. voltage of a photocell at different radiation fluxes.

Obviously it is easier to construct cells for UV or violet wavelengths. However, there are a number of photocathodes enabling one to register radiation in the yellow or red part of the visible range (Cs–Sb deposited on lime glass or quartz) or even radiation in the near infrared (Cs–O–Ag cathode deposited on polished lime glass, for instance).

The quantum efficiency of photoelectric cells is usually in the range of 10 to 25%. The main noise mechanism is evidently the shot noise. A typical graph of cell current vs. voltage supplied is shown in Fig. 4.4. As is evident from this graph there is a linear zone which is followed by a zone of saturation (all generated electrons are participating in the cell current and a further increase of voltage cannot pull more electrons to the anode). The higher the radiation intensity (the flux Φ) the larger the number of electrons and the higher the saturation current in the cell.

Photomultipliers

The main disadvantage of a single photocell is its low-level photocurrent. The situation can be improved drastically if one adds to the cell a process of electron multiplication. This process is based on acceleration of the photoelectrons by an appropriately adjusted electric field and conversion of the excess kinetic energy in secondary emitted electrons. The secondary electrons are further accelerated and cause more electrons to appear, and so on, until the required level of output current is achieved. The number of repeated stages of amplification is usually 8 to 12, and all of them are embedded in a single device, together with the primary photocathode section, comprising the complete photomultiplier tube. The electrodes emitting the secondary electrons are called dynodes and obviously the voltage drop should be set between each two adjacent pairs of dynodes.

A schematic of a typical photomultiplier is depicted in Fig. 4.5. The photocathode is made of a material of appropriate quantum efficiency and is followed by



FIGURE 4.5 Schematic of (a) a photomultiplier and (b) the voltage supply to the dynodes.

the electron focusing and acceleration section and then by the section of dynodes, ending with the anode electrode. The output signal is created as a voltage on a load resistor R_L . The main parameter of the photomultiplier is its total gain, $G_{tot} = G^n$, where *n* is the number of dynodes and *G* is the gain of a single dynode. The anode current is related to the photocathode current as

$$i_{a} = i_{\text{ph.c.}} G^{n}. \tag{4.17}$$

The dominant noise in the photomultiplier is shot noise. It can be shown that fluctuations of the anode current can be described by a formula similar to Eq. (4.11):

$$\overline{i_{a.Sn}^2} = 2ei_a \times \Delta f \times G^{n+1}/(G-1)$$
(4.18)

and consequently the signal-to-noise ratio at the output is lower than that of the photocathode:

$$SNR_{PhM} = SNR_{Phc} \times \frac{G-1}{G}$$
 (4.19)

The rise time and fall time of the photomultiplier signal are very small, so that these devices enable one to handle information at a rate of up to 100 MHz. As to the saturation level of incident radiation, one should keep in mind that usually photomultiplier electrodes, cathode or dynodes or anode, work properly if the current density does not exceed 100–150 nA/cm².

Semiconductor Detectors

With regard to the primary processes in the material, semiconductor detectors act in a way opposite to that which occurs in laser diodes and LEDs as described in Chapter 3. Namely, they convert directly radiation energy into an electric current generated in the semiconductor material. Energy diagrams of typical semiconductors exploited for detecting radiation are depicted in Fig. 4.6. In a pure



FIGURE 4.6 Energy diagrams of (a) intrinsic and (b) extrinsic semiconductor detector materials.

semiconductor substance (intrinsic case) charge carriers (electrons) become free while absorbing energy of incident photons if this energy is enough to promote the electrons from the valence band to the conduction band. Again, as in the case of photocathodes explained earlier, there is a principal limitation of the photon-to-electron conversion process: the gap of the forbidden zone. The photon energy must be greater than E_g and therefore there exists a limiting (maximum) wavelength at which detection of light occurs in the semiconductor:

$$\lambda_{\max} = \frac{1.240}{E_{\rm g}} \tag{4.20}$$

where the wavelength is in μ m and E_g is in eV. For Ge, for example, the gap is 0.66 eV and for Si it is 1.09 eV, both semiconductors evidently being suitable for IR radiation detection as well as for visible radiation detection.

In extrinsic semiconductors an additional energy level might occur inside the forbidden zone as a result of impurities inserted (deliberately) in the crystal lattice. This level can be close to the conduction band or close to the valence band. The first case is referred to as a donor level and the second one as an acceptor level. The donor level might lose electrons which accept additional energy from the incident photons and jump into the conduction zone yielding an excess of free negative carriers (this case is presented in Fig. 4.6b). The acceptor level might receive electrons from the valence band, again as a result of absorbing photons, and then a lack of electrons (holes) is created in the substance. Due to internal processes in the material these holes move like electrons, but in the opposite direction, as an external electric field is applied. These two kinds of extrinsic semiconductors are commonly addressed as n-type (negative carriers) and p-type (positive carriers). Examples of the n-type are Cd–S ($E_g = 2.4 \text{ eV}$) and Cd–Se ($E_g = 1.8 \text{ eV}$) and of the p-type are Ge:Hg and Ge:Cd. If two semiconductor materials, one of



FIGURE 4.7 Schematic of a photoconductive detector.

n-type and the other of p-type, are put in contact with each other a p-n junction is created where carriers of both types are generated when the junction is exposed to radiation of an appropriate wavelength. These p-n junctions are widely exploited in semiconductor detectors.

A great variety of architectures have been developed for semiconductor devices. We will mention here two main groups: photoconductive detectors and photovoltaic detectors. A schematic of the first type is presented in Fig. 4.7. Electric current in a circuit comprising a source of DC voltage, V, and a load resistance, R_L , connected in sequence with the detector of resistance R_d , is affected by the electrons liberated into the conduction band by incident photons. The short current of the detector illuminated by radiation of power P_{λ} is determined as follows (see Keyes, 1977):

$$i_{\rm Sc} = P_{\lambda} \frac{\lambda e}{hc} \eta \mu \tau \frac{V}{a^2} \tag{4.21}$$

where η is the quantum efficiency of the detector material, τ and μ are the carrier lifetime and mobility, and *a* is the size of the detector in the direction of current propagation. Assuming $R_d \gg R_L$ one can get from Eq. (4.21) the following expression for the output signal (voltage on the load resistor):

$$V_{\rm L} = P_{\lambda} \frac{\lambda}{hc} \eta \tau \frac{V}{abtn} \tag{4.22}$$

where a, b, and t are dimensions of the detector and n is the volume concentration of the charge carriers.

Photovoltaic detectors (also called the photodiodes) consist of a p–n junction which actually creates the electric field and excess of moving carriers while it is exposed to the incoming radiation. Hence, in general, such a detector is capable of generating an output signal without being connected to an external voltage source



FIGURE 4.8 Schematic of a photovoltaic detector: (a) without an external voltage source; (b) with a load resistor; (c) typical current vs. voltage characteristics.

(see Fig. 4.8a). In practice, however, a circuit with a voltage source and a load resistor is often exploited (Fig. 4.8b). Moreover, usually photodiodes are operated at a negative voltage bias. On a photocurrent vs. voltage graph, like that shown in Fig. 4.8c, the horizontal section defines the operating range and each curve is related to a corresponding radiation flux.

Both shot noise and thermal noise are experienced in semiconductor detectors and frequently the working range is chosen around the point where the r.m.s. values of both noises become equal. As to the time characteristics, these detectors can be operated at high frequencies up to hundreds of MHz and even sometimes in the GHz range.

Thermal Detectors

In thermal detectors there is no generation of photoelectrons. Instead they are based on the increase of resistance with temperature resulting from absorption of incident radiation. Such a detector, called a bolometer, is shown schematically in Fig. 4.9. The key element of the device is a thin layer (B_1) of a conductor material with a



FIGURE 4.9 Schematic of (a) a bolometer and (b) a Winston bridge with a two-layer detector.

high temperature dependence of resistance and an as low as possible heat capacity. This layer is usually deposited on a small glass substrate positioned in a vessel transparent to the measured radiation. High sensitivity to temperature variation can be achieved if layer B_1 is appropriately designed. However, the selectivity of a single-layer device is very poor, since any variation of temperature, both as a result of incoming radiation and of changes of the surrounding temperature, will cause a change of the output signal. To overcome this shortcoming usually two identical layers, B_1 and B_2 , are deposited on two opposite sides of the glass substrate and both are connected to a Winston bridge (see Fig. 4.9b). Since only the first layer is exposed to incoming photons while any other fluctuation affects both layers, the measured signal originates from the incoming radiation only (obviously if the bridge is initially set to zero).

Bolometers enable one to detect radiation powers of as low as 10^{-10} W. The main advantage of the device, however, is its ability to operate in very wide spectral range, usually from the visible to far infrared (up to $20-25 \ \mu$ m), this range being limited by the transparency of the housing input window.

Problems

4.8. A photomultiplier comprises a photocathode made of K–Cs–Sb (quantum efficiency of 25%) and eight stages of secondary amplification, each with a single dynode of 2.5 gain. The device is designed for measurement of radiation of 10^{-9} W at 400 nm wavelength. Assuming the circuitry is operated at room temperature, find the optimum load resistance of the device.

4.9. A photomultiplier cathode is of 2 cm in size and 50 mA/W responsivity. Assuming that the maximum permissible photocurrent density is 120 nA/cm^2 , find the saturation optical power.

4.10. Find the maximum voltage drop and the maximum dissipated electric power on a load resistor of 10 k Ω incorporated in the output circuitry of a detector with the following performance parameters: $\eta = 0.17$ for $\lambda = 400$ nm; NEP = 10^{-15} W/Hz^{1/2}; DR = 10^{11} ; $\Delta f = 10$ kHz.

4.11. Two detectors made of the same optically sensitive material, but of different active size, $A_1 = 10A_2$, are examined for some application. The optical system is corrected according to the size of the detector, so that the whole light spot of incident radiation is concentrated on the active area of the detector in either case. What benefit, if any, can be gained by using one of the detectors rather than the other?

4.4. Detector Arrays (One-dimensional Arrays and CCD and CMOS Area Sensors)

Simple Arrays

There are several reasons why embedding a number of sensors in a single housing might be attractive for many applications. One of the main reasons is the ability to analyze the spatial distribution of incoming radiation without having to move the detector. The simplest case is a two-element detector (see Fig. 4.10a) allowing one to find the center of a light spot in the direction OX. Another example is a four-quadrant detector (Fig. 4.10b) capable of finding the center of a light spot in both the OX and OY directions. This is done by registering and comparing first the A+B signals vs. C+D signals and then the A+C signals vs. B+D signals. Due to the simplicity of signal processing this type of detector was realized first in analog electronic circuitry and such a configuration was exploited for many years in various optical navigation systems.

New features arise if more than two elements in one line are configured. In this case advanced signal processing might be applied allowing one to find the characteristic points of a light spot (a maximum or a median of the spot intensity distribution), providing the uncertainty is smaller than the pitch p of the array of elements (see Fig. 4.10c; details of this approach are explained in Problems 4.16 and 4.17). Usually a multi-line detector is composed of a number of photodiodes separated mechanically (by grooves) on a common substrate, each one having separate wires for voltage supply and signal output. At present detectors are commercially available with 8, 16, 32, to 128 elements. Evidently a disadvantage of such an array is the great number of wires to be handled. This problem is solved by applying the technology of charge coupled devices (CCDs).



FIGURE 4.10 Simple detector arrays: (a) two-element detector; (b) four-quadrant detector; (c) multi-element line detector.

CCD Detectors

A CCD is an integrated circuit (chip) built of a silicon substrate above which a number of polysilicon transparent electrodes are located (Fig. 4.11a). The electrodes, isolated from the substrate by a SiO₂ layer, are divided in several groups (three in the figure), each group being connected to a separate wire having one of three electric potentials, Φ_1 , Φ_2 , or Φ_3 . Photons of the incident radiation travel through the electrodes and are absorbed in the upper part of the substrate generating photoelectrons. During the time, τ_{exp} , to which the chip is exposed to radiation the photoelectrons are collected in the vicinity of the electrodes where the electric field creates potential wells (shown by dotted lines in Fig. 4.11a). As the exposure time is ended a fast read-out procedure begins ($\tau_{read} \ll \tau_{exp}$) during which the potentials Φ_1 , Φ_2 , and Φ_3 vary in such a way that the electrons collected under each electrode are transferred (shifted) in a three-step process to the adjacent element, all together as one block, as depicted in Fig. 4.11b where potential wells in three sequential time intervals, t_1 , t_2 , and t_3 , are shown.

The variation of the wire potentials is then repeated, pushing the electrons further along the array, until they finally come to the output diode and are read out to the external electronic circuit. Thus, at the output of the CCD arrangement photoelectrons are emerging as charge pulses, sequentially, one by one, through a single wire, no matter how many elements there are in the array. The charge values represent the spatial distribution of light intensity along the CCD line (Fig. 4.12).

The type of detector discussed above is a one-dimensional (1-D) array. Further development of the CCD approach results in two-dimensional (2-D) arrays widely exploited as area sensors capable of capturing a full image in a single shot. A variety of possible architectures have been implemented: one of them is shown schematically in Fig. 4.13a. The image area is a 2-D array of elements (picture elements, or



FIGURE 4.11 CCD detector: (a) schematic of a basic configuration; (b) potential wells and charge transfer.



pixels), each one is like that of Fig. 4.11a and altogether they are arranged in rows and columns connected horizontally to wires having potentials ΦV_1 , ΦV_2 , or ΦV_3 (vertical shifting) and vertically to the elements of the line readout shift register (horizontal shifting guided by potentials ΦH_1 , ΦH_2 , and ΦH_3). As in the case of a line array, the sensor is first exposed to incident radiation (exposure time) followed by a read-out procedure governed by the timing of switched vertical and horizontal potentials. That is, first the lowest horizontal line is transferred at once to the line shift register and read out there, then all the lines go vertically down one step and the next line (initially second from the bottom) is read out through the same line shift register, and so on, until all the horizontal lines are read out in sequence. Again, the exposure time should be much greater than the read-out time, in order to minimize additional noise caused by incident photons while the previously collected charges are still on the chip.

Modern technology allows one to manufacture CCD chips with a tremendous number of pixels (usually hundreds of thousands, but in some cases, like in highresolution digital cameras, up to several millions). Special measures should be undertaken in order to handle so many output charge pulses, to relate properly each one to a corresponding pixel of the CCD matrix, and to reveal in such a manner the incoming image incident on the chip during the exposure time period. The way this is done is to convert the CCD output to a standard video signal exploited for many years in TV engineering and communication. This becomes even clearer if one keep in mind that our final goal is to reverse the output electric signals of the CCD into variation of brightness and represent them using a standard display device, either a video monitor or a computer terminal. A CCD area sensor followed by electronic circuitry where conversion of CCD pulses into a video waveform is carried out in real time constitutes the complete device called a video CCD camera. We will mention here only a few features of video signals which are important for applications from the optical point of view. More details can be found, for example, in Inglis (1993).



FIGURE 4.13 (a) 2-D CCD array and (b) output video signal.

There are several commonly used standards for video signals. All of them are based on the main approach that a display device performs a scanning of the screen area of the monitor, line by line, creating a raster, and the incoming video signals have built-in information at the beginning and ending of each line (synchronizing pulses, or simply "sync"). Near each sync there is a blanking pulse during which the display tube (screen) is darkened and the electron beam of the tube jumps to the starting point of the next line. Thus, the useful information on each line (l_1 , l_2 , l_3 , etc.) is between the sync and blanking pulses of two adjacent lines, as shown in Fig. 4.13b. It should also be taken into account that the black level (BL) signals are usually represented by a higher voltage than the white level (WL) (maximum illumination).

The total number of lines corresponding to a full screen is either 525 (USA standard) or 625 (European standard). However, in order to reduce the bandwidth requirements of the monitor the full image is displayed in two parts, first all even lines and second all odd lines. Hence, the full image signal (called the frame) comprises two half-image signals called the first and the second fields (even lines and odd lines, respectively). This method of image display is referred to as interlaced images. In the last few years, however, the bandwidth available has noticeably improved and CCD video cameras outputting non-interlaced images (called also progressive scan) have become more popular. In any case the standard video frame rate is 30 frames/s for American standard ($\tau_{exp} \approx 40$ ms).

Returning to CCD sensor features, the spectral responsivity of most CCDs is from 400 nm to 1,100 nm, with a maximum at 600 to 850 nm. To keep the CCD spectral response as close as possible to that of the human eye an IR cut-off filter is commonly inserted in front of the CCD chip, reducing the overall sensitivity to IR wavelengths almost to zero. The saturation level of most CCDs is about $0.2 \,\mu$ J/cm². For higher illumination levels the number of electrons generated under a single electrode (pixel) exceeds the ability of the potential well to retain them locally and the photoelectrons start traveling along the silicon substrate interfering with the charge transfer process. However, there are commercially available CCDs designed specially for high dynamic range and their potential well is saturated by 300,000 or even 600,000 electrons. On the other hand, low-intensity signals are limited by dark current (for long exposure) and by read-out noise. Both can be effectively reduced by cooling the CCD chip (see also Section 4.2). With no special adaptations the dynamic range of a CCD is about 200–300. The spatial resolution is dictated by the pixel size. For 1-D arrays a pixel size as small as 3-5 μ m is not unusual. 2-D arrays vary significantly in chip size (from 1/6" up to 2'') as well as in single pixel size (from 5 to 50 μ m).

It should also be mentioned that some area matrix sensors are based on photodiode arrays (CCPD). They are usually faster than CCDs with potential wells and enable one to get a higher frame rate (up to 1,000 frames/s). Arrays of CMOS elements are also available and are becoming more and more popular. They are also based on photodiode elements, but have a higher level of integration than CCPDs or CCDs and therefore can be more compact in size. However, the noise level of CMOS sensors is higher and the sensitivity is lower than that of CCD sensors.

2-D detector arrays intended for color imaging have special features with regards both opto-mechanical architecture and video signal formats. These are discussed in Chapter 10.

Problems

4.12. Optical tracking. A robot equipped with a four-quadrant detector for navigation performs tracking of a target object. At time t_1 the readings of the quadrants are $S_A^{(1)} = 10$; $S_B^{(1)} = 60$; $S_C^{(1)} = 5$; $S_D^{(1)} = 20$ (in relative units). Then the motion correction system is activated and the new readings (at t_2) become $S_A^{(2)} = 15$; $S_B^{(2)} = 40$; $S_C^{(2)} = 65$; $S_D^{(2)} = 70$. Is correction carried out properly?

4.13. If two CCD area sensors having the same number of pixels and working at the same video standard rate but of different size (e.g., the first with a 1/3'' chip and the second with a 2/3'' chip) are available for some application, is there any

advantage of one of them over the other? If yes, what performance parameter is affected by this choice?

4.14. An optical system of 4 cm entrance pupil and 50 mm focal length creates an image of an object plane 10 m distant from the system onto a CCD area sensor of 7 × 7 μ m pixel size, 25% quantum efficiency, 100 electrons read-out noise, and saturation exposure $E_{\text{sat}} = 0.2 \,\mu$ J/cm². Assuming that the reflectivity of the object plane is R = 0.6 and the transparency of optics T = 90%, find the minimum illumination level required for proper differentiation of objects at minimum contrast C = 5%.

4.15. An optical system creates an image of 1.2 mm field of view on a CCD area sensor at magnification V = -10. The required resolution of imaging is 200 lp/mm (in the object plane). Find the minimum required number of pixels in each line of the CCD and the size of a single pixel.

4.16. *Image location with sub-pixel accuracy.* A line CCD detector built of 10 μ m pixels is located in the output plane of a spectrometer. The spot of a spectral line captures three sequential pixels and the corresponding readings are $S_1 = 33$; $S_2 = 127$; $S_3 = 80$ in relative units (called gray levels; they are integer numbers resulting from analog-to-digital conversion and digitization of the CCD output signals). Assuming the light distribution inside the spot to be a symmetrical function: (a) find the location of the spot relative to the center of the second pixel; (b) how precise is the result if SNR of the CCD output is 30:1?

4.17. Sub-pixel accuracy in 2-D space. A light spot is incident on a CCD area sensor built of pixels of $10 \times 10 \,\mu\text{m}$ in size and captures a 3×3 pixels area. The corresponding readings of the sensor pixels constitute the following matrix:

20	110	57
30	150	80
12	70	35

Assuming the spot is symmetrical in both the OX and OY directions, find the location of the spot center with regard to the center of pixel S_{00} where the reading is 150 relative units.

4.5. Solutions to Problems

4.1. Substituting the wavelength and the quantum efficiency in Eq. (4.1) yields

$$R = \frac{\eta \times \lambda}{1.24} = \frac{0.1 \times 0.83}{1.24} = 66.9 \text{ mA/W}.$$

4.2. As the figures of merit of both detectors are equal, we get from Eq. (4.4)

$$D^* = \frac{\sqrt{A_1 \times \Delta f}}{\text{NEP}_1} = \frac{\sqrt{A_2 \times \Delta f}}{\text{NEP}_2}; \text{ and } A_1 = A_2 \left(\frac{\text{NEP}_1}{\text{NEP}_2}\right)^2$$
$$= A_2 \left(\frac{2 \times 10^{-14}}{4 \times 10^{-15}}\right) = 25A_2.$$

Denoting the level of illumination as E we obtain from Eq. (4.1) between responsivity and quantum efficiency:

$$\frac{i_{\det 1}}{i_{\det 2}} = \frac{EA_1R_1}{EA_2R_2} = \frac{A_1}{A_2} \times \frac{\eta_1}{\eta_2} = 25\frac{0.3}{0.5} = 15.$$

Therefore, using the first detector will cause 15 times greater current under the same illumination conditions.

4.3. Keeping in mind the charge of a single electron $e = 1.6 \times 10^{-19} C$ we calculate the average noise current as $\bar{i}_n = 2 \times 10^6 \times 1.6 \times 10^{-19}/10^{-3} = 0.32$ nA. From Eq. (4.1) we get the responsivity of the detector $R = (0.86 \times 0.83)/1.24 = 0.576$ A/W which enables one to find the NEP and detectivity: NEP $= \bar{i}_n/R = 0.32 \times 10^{-9}/0.576 = 0.556 \times 10^{-9}$ W; $D = 1/\text{NEP} = 1.8 \times 10^9$ W⁻¹. Finally, by substituting the calculated values in Eq. (4.4) we obtain:

$$D^* = D\sqrt{A \times \Delta f} = 1.8 \times 10^9 \sqrt{3 \times 10^{-2} \times 10^5} = 9.86 \times 10^{10} \text{ cmHz}^{1/2} \text{ W}^{-1}.$$

4.4. If a constant optical power P is incident on the detector active area its current finally will achieve the value PR amperes (where R is the responsivity of the detector). However, due to electrical inertia of the detector circuitry the current increases from zero to the maximum level gradually and this gradual growth can be described as

$$i_{\text{out}}(\tau) = RP\left[1 - \exp\left(-\frac{\tau}{t_0}\right)\right].$$

Since the rise time is defined as the time duration required for the detector current to rise to 90% of the maximum value, we get $RP[1 - \exp(-\tau_{\text{rise}}/t_0)] = 0.9RP$; $\tau_{\text{rise}} = t_0(-\ln 0.1) = 2.3 \times 0.5 = 1.15$ ns. Also, $\tau_{\text{Fall}} = 2.5 \times \tau_{\text{rise}} = 2.645$ ns. Hence, the maximum working frequency (and the bandwidth) of the circuitry responding to the pulse of infinitesimal duration $\delta \rightarrow 0$ is

$$f_{\text{max}} = \frac{1}{2(\tau_{\text{rise}} + \tau_{\text{Fall}})} = \frac{10^9}{2(1.15 + 2.645)} = 124 \text{ MHz}.$$

4.5. (a) For 50 MHz maximum frequency the corresponding sampling time is

$$\tau = \frac{1}{2\Delta f} = \frac{10^6}{2 \times 50} = 10^{-8} \text{ s.}$$

Keeping in mind that a single photon in the visible ($\lambda = 0.5 \,\mu$ m) has energy $hc/\lambda = 4 \times 10^{-19}$ J we calculate the average number of photons coming to the detector at time τ at minimum illumination level as follows: $\overline{N} = \text{NEP} \times \tau/(4 \times 10^{-19}) = 25$ photons.

(b) By substituting this value in the Poisson distribution (Eq. (4.6)) we find the corresponding probability:

$$P(\overline{N}) = \frac{25^{25}e^{-25}}{25!} = 7.95 \times 10^{-2} \approx 8\%.$$

A fluctuation of 20% from the average value means that the number of photons may be reduced to as low as N = 20 and the same expression (Eq. (4.6)) in this case yields

$$P(20) = \frac{25^{20}e^{-25}}{20!} = 5.19 \times 10^{-2} = 0.65P(\overline{N}).$$

Assuming the fluctuations in the range $(+/-)P_{\text{max}}/e = (+/-)0.37P_{\text{max}}$ around the maximum value P_{max} occur frequently enough to be observed, we draw the conclusion that 20% fluctuation in our case (i.e., variation of the number of photons from 25 to 20) will appear in a reasonable time interval.

4.6. Expression (4.14) allows for the calculation of the dark current at both temperatures:

 $T_{1} = 25^{\circ}C = 298 \text{ K}; \qquad i_{d.c.}^{(1)} = \text{const} \times \exp\left(-\frac{1.1 \times 1.6 \times 10^{-19}}{2 \times 1.386 \times 10^{-23} \times 298}\right)$ $= \text{const} \times \exp(-21.306)$ $T_{2} = 0^{\circ}C = 273 \text{ K}; \qquad i_{d.c.}^{(2)} = \text{const} \times \exp\left(-\frac{1.1 \times 1.6 \times 10^{-19}}{2 \times 1.386 \times 10^{-23} \times 273}\right)$

$$= \text{const} \times \exp(-23.257)$$

and therefore

$$\frac{i_{\rm dc}^{(1)}}{i_{\rm dc}^{(2)}} = \exp(1.951) = 7.04.$$

This means that cooling by 25°C causes the dark current to decrease by a factor of about 7.

4.7. Saturation exposure multiplied by the pixel size and divided by the exposure time of a single frame (1/30 s) gives the saturation optical power:

$$P_{\text{sat}} = \frac{0.2 \times 10^{-6} \times 0.49 \times 10^{-6}}{1/30} = 3 \times 10^{-14} \text{ W/px.}$$

Or in photons per pixel

$$P_{\text{sat}} = \frac{3 \times 10^{-14}}{4 \times 10^{-19}} = 75 \times 10^3 \text{ photons/s/px}$$

(taking into account that a single photon of 0.5 μ m wavelength has an energy of 4×10^{-19} J). Furthermore, the maximum current from a single pixel is $i_{\text{max}} = P_{\text{sat}}\eta = 15,000$ electrons/px and therefore the dynamic range can be found from Eq. (4.5) as follows: DR = $i_{\text{max}}/i_n = 15,000/50 = 300$.

4.8. We start with the calculation of the responsivity of the cathode from Eq. (4.1):

$$R = \frac{\eta \lambda}{1.24} = \frac{0.25 \times 0.4}{1.24} = 0.0774 \text{ A/W}$$

and proceed further to the cathode current: $i_{\text{cth}} = R \times 10^{-9} = 7.74 \times 10^{-11}$ A. Since the total gain is $G_{\text{tot}} = 2.5^8 = 1,526$, the current on the anode found from Eq. (4.17) is $i_{\text{an}} = 7.74 \times 10^{-11} \times 1,526 = 1.18 \times 10^{-7}$ A.

We define the optimal point as that where the mean values of the shot noise current and the Johnson noise current become equal, meaning $\bar{i}_{Sn} = \sqrt{2ei_{an}\Delta f} = \bar{i}_{Tn} = \sqrt{4kTR_L\Delta f}$, which yields

$$R_{\rm L} = \frac{4kT}{2ei_{\rm an}} = \frac{4 \times 1.381 \times 10^{-23} \times 300}{2 \times 1.6 \times 10^{-19} \times 1.18 \times 10^{-7}} = 439 \,\rm k\Omega.$$

4.9. The maximum current from the whole area of the cathode is

$$i_{\text{max}} = 120 \times 10^{-9} \times \frac{\pi \times 2^2}{4} = 377 \text{ nA.}$$

This current is caused by radiation of the following power: $P = 377 \times 10^{-9} / (50 \times 10^{-3}) = 7.54 \,\mu\text{W}.$

4.10. Since the responsivity of the detector is $R = (0.17 \times 0.4)/1.24 = 54.8 \text{ mA/W}$ and the maximum acceptable radiation power is $P_{\text{max}} = \text{DR} \times \text{NEP} \times \Delta f^{1/2} = 10^{11} \times 10^{-15} \times 10^2 = 10^{-2} \text{ W}$, the corresponding current generated by the detector will be $i_{\text{max}} = P_{\text{max}}R = 10^{-2} \times 54.8 \times 10^{-3} = 548 \mu\text{A}$. Then the voltage drop on the load resistor is 5.48 V and dissipated electric power is estimated as $P_{\text{el}} = i_{\text{max}} \times V_{\text{L}} = 548 \times 10^{-6} \times 5.48 = 3 \text{ mW}.$ **4.11.** Assuming the shot noise is the dominant noise of the detector circuit and taking into account that the dark current is proportional to the square root of the active area, $i_{d.c.} \propto \sqrt{A}$, on one hand, and that the useful generated current i_s is the same for both cases (all energy is captured by the active area), on the other hand, one can draw to the conclusion that replacing a smaller detector by a larger one will cause the mean fluctuation to increase, as follows from Eq. (4.11): $\sqrt{\overline{i}_n^2} = \sqrt{2ei_d\Delta f} = \sqrt{2e(i_s + i_{d.c.})\Delta f}$. Hence, the signal-to-noise ratio will be higher for a smaller detector:

$$SNR = \frac{i_{s}}{\sqrt{\overline{i}_{n}^{2}}} = \frac{i_{s}}{\sqrt{2e(i_{s} + i_{d.c.})\Delta f}}.$$

Therefore, the smaller the detector the higher the selectivity of the system, i.e., the ability to differentiate between two cases of close (but still different) radiation intensity.

4.12. We describe the current position of the tracking object (the target) by the vector $\overrightarrow{T}(T_X, T_Y)$ with two components referred to the coordinate system with origin in the center of the detector (see Fig. 4.14). Components of this vector are related to the readings of the detector quadrants as follows:

$$T_X = \frac{(B+D) - (A+C)}{A+B+C+D}; \quad T_Y = \frac{(A+B) - (C+D)}{A+B+C+D}$$

The goal of the tracking navigation is evidently the zero vector and correction at each step is aimed at reducing the modulus of the vector relative to its previous value.

At time t_1 we get from these formulae $T_X = (80 - 15)/95 = 0.684$; $T_Y = (70 - 25)/95 = 0.474$ and the vector length is 0.832. After correction, at time t_2 , the corresponding values are $T_X = (110 - 80)/190 = 0.158$; $T_Y = (55 - 135)/190 = -0.421$; $|\vec{T}| = 0.450$. As we see, the target vector is indeed closer to the origin, meaning that the correction works properly.



FIGURE 4.14 Problem 4.12 – The target vector plane.



FIGURE 4.15 Problem 4.14 – Low-contrast imaging with a CCD.

4.13. If the total number of pixels remains the same for both CCDs the size of a single pixel is greater for a bigger chip. Therefore a larger number of photoelectrons can be collected in the pixel potential well, meaning that the saturation level is also increased. Thus, using the CCD with the larger chip allows for operation at higher illumination level. Furthermore, assuming the read-out noise remains the same in both cases, we may expect the bigger chip is also better with regards dynamic range.

4.14. Let two points, A and B, in the object plane P have slightly different reflectivities R_A and R_B which yield the intensity of the reflected light to be (see Fig. 4.15)

$$I_{\rm A} = E_0 s'_{\rm px} R_{\rm A} \frac{\omega}{2\pi}; \quad I_{\rm B} = E_0 s'_{\rm px} R_{\rm B} \frac{\omega}{2\pi}$$

where E_0 is the illumination level, measured in lx, on the plane P and s'_{px} is the area conjugate with a single pixel of the CCD. The contrast C of the object defined as $C = (I_A - I_B)/(I_A + I_B)$ will cause a corresponding difference in the electric charges of the CCD pixels: $C = (N_{eA} - N_{eB})/(N_{eA} + N_{eB}) = \Delta N_e/2N_e = 0.05$. Since the difference between the number of electrons in two relevant pixels should be greater than the read-out noise, $\Delta N_e = 0.05 \times 2 \times N_e \ge 100$ electrons, we can state that the signal charge on the pixel should be equal to 1,000 electrons at least. To find the corresponding illumination level E_0 we should take into account that: (i) the optical magnification (actually minification) of the imaging optics is $V = S'/S = f'/S = -(50/10^4) = -1/200$; (ii) the active area of a single pixel $s_{px} = 49 \times 10^{-12} \text{ m}^2$; (iii) the solid angle $\omega = \pi (40^2)/(4 \times 10^8) = 12.56 \times 10^{-6} \text{ sr}$; (iv) the exposure time at standard video rate is $\tau_{exp} = 1/30$ s; (v) the conversion factor from photometric units to radiometric units in the visible is K = 683 lm/W (for

simplicity we neglect the spectral dependence of luminous efficacy, see details in Chapter 10); and (vi) a single photon in the visible has an energy of 4×10^{-19} J. Then we get for the number of electrons generated in a single pixel of the CCD by incoming light:

$$N_{\rm e} = \frac{E_0}{K} \times \frac{\omega}{2\pi} \times \frac{s_{\rm px}}{V^2} \times \tau_{\rm exp} \times \frac{\eta RT}{4 \times 10^{-19}} = 1,000$$

which gives

$$E_0 = \frac{10^3 \times 4 \times 10^{-19} \times 0.25 \times 10^{-4} \times 683}{2 \times 10^{-6} \times 49 \times 10^{-12} \times 0.033 \times 0.25 \times 0.6 \times 0.9} = 15.64 \,\mathrm{lx}.$$

What remains to check is that this illumination level does not cause saturation of the CCD. At saturation a single pixel of the CCD receives a number of photons $N_{\text{ph.sat}}$:

$$N_{\text{ph.sat}} = \frac{0.2 \times 10^{-6} \times 49 \times 10^{-8}}{4 \times 10^{-19}} = 2.45 \times 10^5 \text{ photons/px}$$

which creates 61,000 electrons ($\eta = 0.25$), i.e., illumination level E_0 generates less than 2% of the number of electrons at saturation.

4.15. A spatial frequency of 200 lp/mm is equivalent to a periodic object with period $T = 1/200 = 5 \,\mu$ m. In the CCD plane the corresponding period is $T \times V = 50 \,\mu$ m and it should be equal twice the pixel size (Nyquist sampling theorem). Therefore, a single pixel is 25 μ m and the number of pixels in one line of the CCD is

$$N = \frac{1.2 \times 10}{25 \times 10^{-3}} = 480 \text{ px.}$$

4.16. Let the light intensity distribution on the CCD look like that of Fig. 4.16.

(a) We assume that all pixels are of the same size p and that the gap between pixels can be neglected. We also choose the coordinate system XOY with origin in the center of the second pixel and assume that the function F(x) describing the intensity distribution of the incident spot is symmetrical, it is spread over the interval (-r < x < r), and its maximum is at a distance a from the point x = 0. To determine location of the spot one should choose some characteristic point in the spot and find its coordinate. Intuitively such a point might be the point of maximum intensity inside the spot. However, it turns out that better results can be obtained by choosing the median, m, which is defined as the center of symmetry of F(x):

$$\int_{-r}^{m} F(x') \,\mathrm{d}x' = \int_{m}^{r} F(x') \,\mathrm{d}x'.$$



FIGURE 4.16 Problem 4.16 – Light spot incident on three sequential pixels.

If the function is absolutely symmetrical the median, of course, coincides with the point of the maximum, otherwise these points have slightly different coordinates. In the case shown in Fig. 4.16 our goal is to find the point x = a. Keeping in mind that the light intensity incident on the pixel is just averaged over the pixel area and denoting

$$S_1 = \int_{-1.5p}^{0.5p} F(x-a) \, \mathrm{d}x; \quad S_2 = \int_{-0.5p}^{0.5p} F(x-a) \, \mathrm{d}x; \quad S_3 = \int_{0.5p}^{1.5p} F(x-a) \, \mathrm{d}x$$

we have for the median

$$S_1 + \int_{-0.5p}^{a} F(x-a) \, \mathrm{d}x = \int_{a}^{0.5p} F(x-a) \, \mathrm{d}x + S_3$$

or

$$S_1 + S_2 = 2 \int_{a}^{0.5p} F(x - a) \,\mathrm{d}x + S_3. \tag{A}$$

The integral in Eq. (A) can be expressed as follows:

$$I = \int_{a}^{0.5p} F(x-a) \, \mathrm{d}x = \tilde{F}(x-a) \times (0.5p-a).$$
(B)

Expression (B) is a precise one, but the value $\tilde{F}(x - a)$ from the interval [*a*; 0.5*p*] is not known. We substitute it with the value $F^* = S_2/p$, then Eqs. (B) and (A)



FIGURE 4.17 Problem 4.17 – Light spot incident on 3×3 pixels of a 2-D array.

yield $S_1 + S_2 - S_3 = 2S_2(0.5 - a/p)$ and finally

$$\frac{a}{p} = 0.5 \left(\frac{S_3 - S_1}{S_2}\right).$$
 (C)

It should be mentioned that the same approach can be exploited if the spot captures more than three pixels. In such a case instead of Eq. (C) one gets

$$\frac{a}{p} = 0.5 \left(\frac{\Sigma_3 - \Sigma_1}{S_2}\right)$$

where Σ_1 and Σ_3 are the sums of the signals on the left and on the right from the center pixel, respectively. By substituting the data of the problem in Eq. (C) we get $a/p = (80 - 33)/(2 \times 127) = 0.185$ and therefore the spot center is shifted 1.85 µm right from the center of the second pixel.

(b) To estimate the accuracy of Eq. (C) we denote z = a/p and proceed as follows:

$$\frac{\Delta z}{z} = \frac{\Delta (S_3 - S_1)}{S_3 - S_1} + \frac{\Delta S_2}{S_2}; \quad \Delta z = \frac{2\Delta S}{2S_2} + \frac{\Delta S_2}{S_2} z = (1 + z) \frac{\Delta S}{S_2} = (1 + z) \frac{1}{\text{SNR}}.$$
 (D)

In our case we have $\Delta z = (1 + 0.185)/30$; $\Delta a = \Delta z \times p = 0.40 \,\mu\text{m}$.

4.17. We use the same approach as in Problem 4.16 to treat the function F(x, y) of two coordinates and denote the matrix of the relevant pixels as S_{ij} (i = -1; 0; 1, j = -1; 0; 1) (see Fig. 4.17). Let the pixel size in the OX or OY direction be *p*. Aiming to find the segments a_x and a_y we define again the median, as in the 1-D case, but separately for the OX and OY direction. Then for

the location of a_x we have

$$\int_{-1.5p}^{a_x} \int_{-1.5p}^{1.5p} F(x - a_x, y - a_y) \, \mathrm{d}x \, \mathrm{d}y = \int_{a_x}^{1.5p} \int_{-1.5p}^{1.5p} F(x - a_x, y - a_y) \, \mathrm{d}x \, \mathrm{d}y$$

which can be written in the following manner:

$$\sum_{j} S_{-1,j} + \int_{-0.5p}^{a_x} \left\{ \int_{-1.5p}^{-0.5p} F + \int_{-0.5p}^{0.5p} F + \int_{0.5p}^{1.5p} F \right\} = \int_{a_x}^{0.5p} \left\{ \int_{-1.5p}^{-0.5p} F + \int_{-0.5p}^{0.5p} F + \int_{0.5p}^{1.5p} F + \int_{0.5p}^{1.5p} F + \int_{0.5p}^{1.5p} F + \int_{0.5p}^{1.5p} F \right\} + \sum_{j} S_{1,j}.$$
 (A)

Substituting again the integrals of F with the average value of the corresponding pixel:

$$\iint F = S_{i,j}p^2$$

we obtain from Eq. (A):

$$\sum_{j} (S_{-1,j} - S_{1,j}) + \sum_{j} S_{0,j} = \frac{2p(0.5p - a_x)}{p^2} \{S_{0,-1} + S_{0,0} + S_{0,1}\}$$

and finally

$$a_x = \frac{p}{2} \frac{\sum_{j} (S_{1,j} - S_{-1,j})}{\sum_{j} S_{0,j}}.$$
 (B)

The median in the OY direction can be treated in a similar way:

$$a_{y} = \frac{p}{2} \frac{\sum_{i} (S_{i,1} - S_{i,-1})}{\sum_{i} S_{i,0}}.$$
 (C)

Thus, the numerical data of the problem give

$$\frac{a_x}{p} = \frac{(57+80+35) - (20+30+12)}{2(110+150+70)} = 0.166;$$
$$\frac{a_y}{p} = \frac{(20+110+57) - (12+70+35)}{2(30+150+80)} = 0.135$$

and therefore the center of the spot is located 1.66 μ m to the right and 1.35 μ m upwards of the center of the pixel (0,0).

This page intentionally left blank

Optical Systems for Spectral Measurements

5.1. Spectral Properties of Materials and Spectral Instruments

Wavelength-dependent features specific to a material and related to the generation or propagation of electromagnetic radiation are called the spectral properties of the material. There are spectral properties corresponding to emission, absorption, and scattering of electromagnetic waves.

Emission Spectra

The simplest case is spontaneous radiation of a material in the atomic state. It is well known that each kind of atom is characterized by a specific energy diagram (Fig. 5.1a): each horizontal line represents the possible level of energy which the atom, being excited, might possess. Each transition from a higher level (say, 1 or 2) to a lower level (say, 0) is accompanied by emission of a photon of energy:

$$E_1 - E_0 = hv_1; \quad E_2 - E_0 = hv_2$$

and this is represented on the wavelength scale (or optical frequency scale, see Fig. 5.1b) by a corresponding peak (delta-function) called the spectral line. The height of each peak represents the intensity of the radiation, I, at a specific optical frequency. This depends on the probability of the transition between corresponding energy levels (which is described in terms of the number of atoms at each energy level, the lifetime of the atoms at each level, and some other properties of atoms). A graph like that of Fig. 5.1b is called the emission spectrum.



FIGURE 5.1 (a) Energy diagram and (b) emission spectrum.

The main features of the emission spectrum are the location of the spectral lines and their relative intensities. These features are very specific for each kind of atom and this is the main reason why the emission spectrum is widely used for the identification of different atoms present in a compound to be tested. Of course, the energy diagram and the emission spectrum of many atoms are much more complicated than the examples shown in Fig. 5.1. This is also true for ions or molecules where the number of degrees of freedom are significantly higher than in a simple atom. As a result, the energy diagram has many more possible energy levels and the corresponding emission spectrum is rich in spectral lines widely spread not only in the visible, but also in the IR wavelength region.

It should be mentioned that in the numerical and graphical presentation of spectra three types of units are widely used: (i) wavelength λ (usually in micrometers, μ m, or nanometers, nm, or angstroms, Å); (ii) optical frequency $\nu = c/\lambda$ (in Hertz, Hz); (iii) wavenumber $N = 1/\lambda$ (sometimes also denoted as ν ; in cm⁻¹; if λ is in μ m then $N = 10,000/\lambda$ cm⁻¹).

In reality each spectral line is far from being a delta-function (Fig. 5.2). It has a finite width, $\delta\lambda$, and a specific shape, $I(\lambda)$, governed by several physical mechanisms of which we will mention here the following basic three:

- (a) Attenuation of the atom oscillations while emitting the photons. This causes a slight, but finite broadening of energy levels; the corresponding width of the spectral line is called the natural width, $\delta\lambda_n$, and it is estimated by the value 1.2×10^{-4} Å (1 Å = 10^{-8} cm).
- (b) Broadening of spectral lines due to collisions between atoms of a radiating gas, $\delta \lambda_c$. This can be estimated by the expression

$$\delta \nu_{\rm c} = \frac{2\rho^2 p}{\sqrt{2\pi kTm}} \tag{5.1}$$



FIGURE 5.2 Spectral lines of finite width and different shapes.

where ρ is the distance between centers of colliding particles of equivalent mass *m* and *p* is the gas pressure of the atom mixture. Its shape is governed by the formula

$$I_{\nu} = \frac{\delta}{\pi [\delta^2 + (\nu - \nu_0)^2]}$$
(5.2)

where v_0 is the line center frequency and δ is the line width of an infinitely thin radiating layer.

(c) Broadening due to thermal motion of atoms (Doppler broadening, $\delta\lambda_D$). This is described as

$$\delta\lambda_{\rm D} = 7.18 \times 10^{-7} \lambda_0 \sqrt{\frac{T}{M}}$$
(5.3)

with the spectral line shape as

$$I_{\lambda} = I_0 \exp\left[-\left(\frac{\lambda - \lambda_0}{\delta \lambda_{\rm D}}\right)^2\right]$$
(5.4)

where T is the absolute temperature and M is the atomic or molecular weight. Usually the Doppler width is much more significant than the other broadening factors, especially at high temperatures (see Problems 5.3-5.5).

In practice, in order to observe the emission spectrum of a material (e.g., a metal alloy) a sample of it (a slab or a rod) is introduced into an electrical discharge arc where, due to the high temperature, the solid alloy is disassembled into separate atoms and ions which are energized and start to emit radiation. The arc with the sample is used as the radiation source positioned at (or projected to) the entrance plane of a spectral instrument and the emission spectrum of the compound is created and analyzed in the output plane of the device. Usually an entrance slit is positioned in the entrance plane of the spectral device. It is the images of this slit



FIGURE 5.3 Example of an emission spectrum.

appearing at separate locations corresponding to each active wavelength that create the emission spectrum in the output plane. An example of such a spectrum created at the exit of a spectrometer is shown in Fig. 5.3. Different wavelengths appear at different positions in the horizontal direction. If the spectrum is photographed on a film then the intensities of the spectral lines are related to the optical density of the corresponding images on the photograph.

The emission spectra of all chemical elements and many molecules are well known and tabulated, so that the spectral analysis of an unknown compound requires careful comparison of the observed emission spectrum with tabulated data. The concentration of the elements in the compound (which are also usually unknown) evidently affects the relative intensity of the spectral lines and should also be taken into account.

Another example of the usefulness of emission spectra comes from astrophysics. Here valuable information is obtained not only from the position and intensity of spectral lines (which allow one to identify different elements in the atmospheres of stars and planets), but also from analysis of the shapes of spectral lines enabling one to understand different physical processes occurring in the universe.

Absorption Spectra

In the terminology of quantum mechanics absorption of radiation by atoms or molecules is described in a manner very similar to that of emission of electromagnetic waves. It is demonstrated in Fig. 5.4a where absorption of two photons,



FIGURE 5.4 (a) Energy diagram and (b) absorption spectrum.



FIGURE 5.5 Absorption spectrum of solar radiation at the earth's surface.

the energy of which fits exactly transitions between energy levels 0-1 and 0-2, is shown. Fig. 5.4b demonstrates the corresponding change in the intensity distribution of incoming continuous-wavelength radiation from an external source, I_S (say, black body radiation). Obviously, the decrease of intensity of the incoming radiation passing through a mixture of atoms due to absorption phenomena depends on the number of atoms at various energy levels and, therefore, is influenced by the concentration of atoms and the optical path of radiation in the absorbing media.

Absorption spectral lines have in reality a finite width governed by the same conditions and rules as that of emission spectra, as described above.

An example of a real absorption spectrum is presented in Fig. 5.5, which is the spectrum of solar radiation at sea level on earth (after passing through the atmosphere). Absorption lines related to oxygen, water, and CO_2 are clearly identified.

Scattering Spectra

A medium transparent to electromagnetic waves is illuminated by radiation of an external source having a few spectral lines (of frequencies v_1 , v_2 , etc.). The photons of the incident radiation are scattered by molecules of the medium according to the laws of quantum mechanics which take into account not only quantization of energy but also quantization of moments of moving particles (rotation and vibration of molecules and radicals). If the frequencies corresponding to the molecule motion are $v_1^{(m)}, v_2^{(m)}, \ldots, v_i^{(m)}$, etc., the scattered light comprises all possible combinations of incident frequencies with those of the molecules, e.g., in the scattered radiation spectrum new frequencies appear: $v_1' = v_1 - v_1^{(m)}$;



FIGURE 5.6 (a) Raman spectrum of CCl_4 molecules and (b) the spectrum of incident radiation of a mercury lamp.

 $v_2' = v_1 - v_2^{(m)}; \dots v_k' = v_k - v_i^{(m)};$ etc.; and also $v'_{k+1} = v_1 + v_1^{(m)}; v'_{k+2} = v_1 + v_2^{(m)}; \dots v'_{2k} = v_k + v_i^{(m)},$ etc. These two groups of new lines are called the violet and red satellites of the corresponding lines of the incident spectrum and the phenomenon itself is called Raman scattering.

The violet satellites are usually weaker than the red ones, but this difference is reduced with increasing temperature of the scattering media. In general, intensities of the satellite's lines are much lower than those of the incident spectral lines and this leads to difficulties in the observation of Raman spectra in practice. In spite of this, analysis of Raman spectra has become a powerful tool in the study of the molecular structure of materials. This is especially true for complex organic compounds when other (chemical) methods become ineffective or even useless.

An example of a Raman spectrum is shown in Fig. 5.6. The upper part of the figure is the spectrum of the incident radiation and the lower part is the spectrum of scattered radiation. Both kinds of satellites are clearly distinguished.

Luminescence Spectra

Luminescence is defined as the ability of a substance to emit radiation after it is excited by some kind of incident energy, either radiant or non-radiant, providing that the excitation is not thermally originated. Luminescent light is definitely different from thermal radiation – it is governed by different physical laws and conditions from those mentioned in Chapter 6 (e.g., Kirchhoff 's law is valid for thermal radiation of any body but is not applicable to luminescent light). In other words, luminescence is a property related to a medium which is not in a thermal equilibrium state.



FIGURE 5.7 (a) Absorption and (b) luminescence processes.

The creation of luminescence in a substance excited by some incident radiation is demonstrated in Fig. 5.7. A typical energy spectrum of a substance with molecular structure is presented in Fig. 5.7a. The energy levels constitute a number of groups (called the spectrum bands) each one having several lines close to each other. An incident photon of high energy (UV radiation or X-rays) absorbed by the molecules of the substance causes the energy to increase from one of the levels of band 1 to one of the levels of band 3. Another photon of slightly different energy can be also absorbed causing a transition to another energy level of the same band. As a result, an absorption band of a certain width is created (see the lower part of Fig. 5.7a). A small part of the absorbed energy of each incident photon is lost by the molecule (due to mechanical motion or collisions with other particles). Then the molecule comes to the lowest energy level of this spectral band. From here the molecule undergoes a transition to one of the levels of the lower spectral band, 1, while the corresponding photons are emitted. In a collection of molecules all possible transitions are realized and the emitted spectrum (spectrum of luminescence) is a collection of spectral lines typical for given substance (and practically independent of the kind of excitation photons). Obviously the luminescent radiation has greater wavelength (less energetic photons) than the excitation radiation (Stokes' rule):

$$h\nu_{\rm L} < h\nu_{\rm E}$$

(this effect is indicated in the lower part of Fig. 5.7b).

An important characteristic of luminescence is the quantum efficiency of the luminescence process, η_L , defined as the ratio of luminescent energy to the absorbed energy of the excitation photons:

$$\eta_{\rm L} = \frac{E_L}{E_a}.\tag{5.5}$$

If a single photon of high energy (short wavelength, λ_a) causes emission of a number of photons of different wavelengths, λ_i , inside the luminescence spectrum, the following expression is related to the number of photons obtained:

$$\sum_{i} \frac{N_i}{\lambda_i} = \frac{\eta_{\rm L}}{\lambda_a}.$$
(5.6)

Other important characteristics of luminescence are the time delay between absorption and emission of radiation and the duration of luminescence. Longduration luminescence is usually termed phosphorescence and the short-duration process is called fluorescence. The latter is widely used in biology and medicine (e.g., analysis of live cells in fluorescence microscopy or X-ray imagers).

Some important applications (like computed radiography, see Problems 5.9 and 5.10) are based on photostimulated luminescence (PSL). This phenomenon is realized mainly in solids of complex compounds doped with ions of rear earth elements (like BaFBr doped with Eu^{3+} or KBr doped with In^{2+}). In the vicinity of the doping ions additional energy levels with very long lifetime are created. Such areas, called F-centers (see Fig. 5.8), enable the energy of incident high-energy photons to be stored for a long time (level 2), until absorption of additional photons resulting from an external radiation source (like a laser) stimulates transition to a new energy level (3) with very short lifetime so that the further transfer to lower energy levels (4) occurs immediately, accompanied by emission of new photons of luminescent radiation. Usually the stimulated photons possess less energy than the luminescent ones and therefore $\lambda_L < \lambda_{St}$.



FIGURE 5.8 Schematic of photostimulated luminescence (PSL).

Reflectance and Transmittance of Condensed Media

Propagation of electromagnetic waves in solids and liquids as well as transfer of radiation from one media to another are significantly affected by the refractive index of materials (which is actually the main parameter characterizing electromagnetic phenomena in condensed media):

$$n = n_0 - i\chi \tag{5.7}$$

where the real part n_0 determines the group velocity of the waves (related to the speed of the energy transfer and direction of propagation) and the imaginary part χ is related to the decay of radiation due to true absorption in the substance. Both n_0 and χ (often called the optical constants of a material) depend on the wavelength of the propagated radiation. For instance, the refractive index of optical glasses, as pointed out in Chapter 2, varies as $n \propto (\lambda)^{-2}$. In dielectric materials usually $\chi \ll n_0$, except in the regions of strong absorption bands, so that χ can be neglected and $n = n_0$. As a result, refraction in dielectrics obeys simple relations, like the basic formula of refraction Eq. (1.1). Reflectance *R* in this case is governed by Fresnel's formula:

$$R = \frac{1}{2}(R_{\rm s} + R_{\rm p}) = \frac{1}{2} \left[\left(\frac{\sin(i-r)}{\sin(i+r)} \right)^2 + \left(\frac{\tan(i-r)}{\tan(i+r)} \right)^2 \right]$$
(5.8)

where *i* and *r* are the incident and refractive angles (see Fig. 1.2), and R_s and R_p are related to S and P polarization components of the incident radiation. As can be easily shown, for small angles *i* the Fresnel's formula gives

$$R = \frac{(n-1)^2}{(n+1)^2}.$$
(5.9)

In metals and most semiconductors both optical constants are of the same order or magnitude and therefore χ cannot be neglected. In this case reflectance *R* is governed by a much more complicated expression than Eq. (5.8) (see Born and Wolf, 1968, Chapter 13), but for normal incidence (*i* = 0) a simple formula can still be obtained:

$$R = \frac{(n_0 - 1)^2 + \chi^2}{(n_0 + 1)^2 + \chi^2}.$$
(5.9a)

Since n_0 and χ both depend on wavelength, reflectance also is spectrally dependent. The same is true for transmittance, T, defined as the ratio of the intensity of radiation I_d passing through a slab of a given thickness, d, to the intensity at the entrance of the slab, I_0 : $T = I_d/I_0$; and it obeys Bouguer's law:

$$I_d = I_0 \exp(-\alpha d) \tag{5.10}$$


FIGURE 5.9 Absorption of fused silica glass at temperatures of 300 K (1), 700 K (2), and 1,100 K (3).

where the absorption factor, α , is related to the optical constant, χ , as

$$\alpha = \frac{4\pi\,\chi}{\lambda}.\tag{5.11}$$

Dimensions of α are m⁻¹ (or cm⁻¹) and because λ is very small the absorption factor becomes significant even at very low values of χ . For example, if $\chi = 10^{-4}$ for a wavelength of 0.42 μ m the intensity of light is reduced 20 times (to only 5%) after propagation through a slab of material of 1 mm in thickness.

The absorption factor is frequently used in order to describe the spectral behavior of the absorption properties of a substance. An example is shown in Fig. 5.9 where the absorption of fused silica glass in the near IR at different temperatures is presented.

Classification of Spectral Instruments

Although there exists a great variety of spectral instruments they can be classified with regard to: (i) destination; (ii) type of dispersive elements and main architecture; and (iii) ability with regards spectral resolution.

With regard to *destination*, there are monochromators (intended for the creation of monochromatic radiation of a chosen wavelength), spectrometers (intended for registration and measuring the entire spectrum of a sample), and spectrophotometers (intended for measuring transmission and absorption factors of solids and liquids). With regard to the *main architecture* of the instrument, there are devices with prisms and with diffraction gratings, and devices of

interferometric configuration. As to the *spectral resolution ability*, there are systems of low resolution, of high resolution, and of super high resolution. We address in this chapter all classes of instruments.

No matter which dispersive element is exploited in a device or what architecture is chosen, the following parameters serve as basic characteristics of the instrument:

- Angular dispersion of the dispersive element, $d\varphi/d\lambda$ (rad/nm).
- *Linear dispersion* at the exit plane (the spectrum plane), dl/dλ (mm/nm). More frequently the reciprocal value is used (reciprocal liner dispersion): dλ/dl (nm/mm) which is the spectral interval incident on a 1 mm segment of the exit plane.
- Spectral resolution, $\Re = \lambda/\delta\lambda$ (dimensionless), where $\delta\lambda$ is the minimum resolvable spectral interval. For low-resolution devices \Re is usually about $10^3 10^4$ whereas in super high-resolution systems \Re can achieve a value of 10^6 or more.

Problems

5.1. Find the wavenumber and the energy of transition (in eV) corresponding to a radiated green line of $\lambda = 5,460.75$ Å.

5.2. Find the physically limited width (the "natural width") of the spectral line centered at 6,000 Å in terms of wavenumbers and in terms of frequencies.

5.3. The spectrum of iron in the atomic state has a typical triplet (three very close spectral lines) in the UV: 3,100.67 Å, 3,100.31 Å, and 3,099.97 Å. Assuming that iron is present in an arc discharge of 10,000 K, calculate the Doppler broadening of the lines and show that the triplet is still resolvable.

5.4 It is known that in the spectrum of the sun's corona the Fraunhoffer absorption lines are hardly detectable and some of them do not appear at all. Show that this effect can be explained by scattering of the photons emitted by very fast electrons present in the corona. The estimated temperature of the corona's electrons is $T_e = 600,000$ K and Doppler broadening is related to the mass of the electron, m_e , as follows:

$$\Delta v_{\rm D} = \frac{v_0}{c} \sqrt{\frac{2kT_{\rm e}}{m_{\rm e}}}.$$

Perform the calculation for the spectral absorption line at 3,934 Å.

5.5 Find the Doppler width of the main line of a He–Ne laser ($\lambda = 6, 328$ Å) if it originates in the motion of the Ne atoms and the temperature of the laser is 350 K.

5.6. If monochromatic light of wavelength $\lambda = 5,000$ Å is perpendicularly incident on a pure and smooth metallic surface of Au, Ag, Cu, or Ni, what is the percentage of reflected energy in all four cases? Use the optical constants, n_0 , χ , from the following table:

	Au	Ag	Cu	Ni
n_0	0.37	0.18	0.64	1.79
χ	2.82	3.64	2.62	3.32

5.7. In a study of the Raman spectrum of toluene the spectral lines with the following wavenumbers were registered: $3,067 \text{ cm}^{-1}$; $3,054 \text{ cm}^{-1}$; $3,032 \text{ cm}^{-1}$; $2,981 \text{ cm}^{-1}$; $2,920 \text{ cm}^{-1}$; $2,870 \text{ cm}^{-1}$; $1,605 \text{ cm}^{-1}$. Find the location of each line relative to the line of the shortest wavelength in the exit plane of the spectrometer with reciprocal linear dispersion of 50 nm/mm.

5.8. A low-resolution system used for demonstration purposes is exploited in order to demonstrate the ability to separate two close spectral lines, like a typical doublet of cooper (violet doublet) where $\lambda_1 = 4,062$ Å and $\lambda_2 = 4,022$ Å. In the exit plane of the system a line CCD detector array with 10 µm pixel pitch is positioned. Calculate the system resolution and the minimum required linear dispersion.

5.9. *Computed radiography (CR).* In CR for mammography applications (testing of X-ray images for the early detection of breast cancer) a photostimulated luminescent (PSL) effect in a plate made of BaFBr doped with Eu is frequently used. Such a plate, being first exposed to X-rays and then undergoing excitation by a He–Ne laser, has a luminescent line of 390 nm. X-rays in mammography systems are usually of 20 keV (soft X-rays). Assuming that the energetic efficiency of the PSL plate is $\eta_e = 2.5\%$, find its quantum efficiency (defined as the number of generated light photons per single absorbed X-ray photon).

5.10. *System for CR.* The CR approach mentioned in Problem 5.9 can be realized in a number of configurations, one of which is shown in Fig. 5.10. X-rays from a source A are transmitted through a test object B and then are incident on a PSL plate where the rest of the X-ray energy is absorbed and stored as a latent image. Each element of this latent image is actually a collection of F-centers generated in the PSL plate and proportional to the local energy of the incident X-ray beam. Reading out of the latent image is performed sequentially, point-by-point, when an excitation laser beam of wavelength λ_{ex} , after expansion and focusing by the optics SO in a small light spot, causes the F-centers present in the spot area to generate luminescence photons of wavelength λ_L . The luminescence is transmitted through a beam splitter BS and collected by a collection lens L₂ on a photomultiplier tube PhM.



FIGURE 5.10 Problem 5.10 – Configuration of a computed radiography system.

An X–Y scanner SC directs the laser beam to the PSL plate and allows one to get information from all points of the relevant area. The BS has high reflection for λ_{ex} and high transmittance for λ_L (see the graph to the top right of Fig. 5.10). The output signal of the PhM is digitized and processed very fast and stored in a memory buffer, each cell of the buffer corresponding to a separate point of the scanned PSL plate. The contents of the buffer are transferred to a display, creating on the screen a black-and-white pattern of the latent image.

During the read-out process of each spot the number of activated F-centers is reduced until all of them disappear. Suppose that each absorbed photon of λ_{ex} interacts with a single F-center and there is no delay between excitation and luminescence radiation. Then we can state that the dynamic of this process is governed by the simple relation $dN = -\sigma IN dt$ which yields $N = N_0 \exp(-\sigma It)$, where N_0 is the initial number of F-centers, I is the intensity of the laser radiation (in photons/cm²), t is the exposure time, and σ is the absorption cross-section of a single F-center (in cm²). It is evident that the smaller the value N/N_0 the more effective is the read-out process and the higher the PhM signal. On the other hand, there is a limitation of the read-out time which results in limited exposure, t, per single spot (pixel of the final pattern). Assuming that: (i) the X-ray radiation is of 20 keV; (ii) the absorption cross-section of a single F-center, is $\sigma = 10^{-16} \text{ cm}^2$; (iii) the maximum read-out time for the whole PSL plate of size $250 \text{ mm} \times 250 \text{ mm}$ should not exceed 5 min; (iv) the spatial resolution required is 10 lp/mm; and (v) the reflectance of the BS for the laser wavelength is 0.75, find the minimum required laser power in such a system.

[Note: The read-out is satisfactory if up to 90% of F-centers are converted into light photons at each spot.]

5.11. (a) Assuming the shape of the spectral line in the exit plane of a spectrometer obeys the expression

$$I_{\varphi} = I_0 \frac{\sin^2 v}{v^2}$$
, where $v = \frac{\pi}{\lambda} \sin(\varphi - \varphi_0)$

and I_0 and φ_0 are the intensity and the angular location of the center of the line, find the contrast, *C*, in the mutual pattern of two lines, λ and $\lambda + \delta \lambda$, corresponding to the Rayleigh criterion of limiting resolution (minimum resolvable spectral interval $\delta \lambda$) and having equal intensities in their centers.

[Note: Definition of the contrast *C* in this case is $C = 1 - I_{\min}/I_{\max}$; i.e., it differs from the definition of Chapter 2 related to MTF.]

(b) Suppose that the minimum contrast which is still resolvable in the pattern is 5%. How much could one line be weaker than the second one if they are analyzed in the same instrument as in (a)?

5.2. Prism-based Systems

An architecture of a spectral system with a prism as a dispersive element is depicted in Fig. 5.11. Radiation from a light source A (which is analyzed or used for the generation of monochromatic light) is concentrated by the illumination optics (elliptical mirror M in this example) onto an entrance slit S positioned in the front focal plane of a collimator objective lens L_1 . The parallel beam incident on the prism P, is dispersed by it (separated according to the wavelength) while passing through and the output monochromatic beams are focused by an output objective L_2 in the plane T where an output slit S' (monochromator) or output detector array (spectrometer) are located. The slits S, S' are special diaphragms of several millimeters in length and of very small variable width (precisely set, at sub-micrometer accuracy).

If the system works as a monochromator and slit S' is positioned permanently, the prism should be rotated in order to bring different wavelengths to the output.



FIGURE 5.11 Basic configuration of a prism-based spectrometer.

Alternatively, the prism might be fixed and slit S' moved along the T plane. If the system is arranged as a spectrometer slit S' is removed and the whole spectrum is created simultaneously on the elements of the output detector array (usually a line CCD detector, like those described in Section 4.4).

The prism is made of a transparent material with dispersion power $dn/d\lambda$ and is usually oriented in such a way that the medium working wavelength corresponds to the angle of minimum deviation (see Problem 1.17 for details). The refraction angle β of the prism affects the angle ψ between the incident and the output beams in the following manner:

$$\psi = 180^{\circ} + \beta - 2\sin^{-1}\left(n\sin\frac{\beta}{2}\right).$$
 (5.12)

We now consider the minimum resolvable spectral interval and optimal width of the entrance slit. The spectral resolution of the instrument is the main feature of the system. This is defined, as we saw, by the minimum resolvable spectral interval, $\delta\lambda$, which is determined according to the Rayleigh criterion. Namely, two wavelengths, λ and $\lambda + \delta\lambda$, can be still resolved (i.e., registered separately) if their intensity distributions in the output plane correspond to the graph of Fig. 5.12 (the distance between the central points is at least as small as half of the spectral line width).

Obviously the resolution is strongly affected by the width and shape of the spectral line. Both are limited by several physical phenomena, as explained in the previous section. However, they are also dependent on the instrument parameters, in particular on the width of the entrance slit. Besides this, the illumination conditions at the entrance of the device as well as aberrations of the imaging optics are also important.

The optics of the device creates a geometrical image of the entrance slit of width *s* in the output plane. For ideal optics the size of such an image obeys the expression



FIGURE 5.12 Rayleigh's criterion of spectral resolution.

where f'_1 and f'_2 are the focal lengths of the collimator and output objectives, respectively. However, the true width of the line in the plane T is also governed by: (i) aberrations of the lenses and the prism; (ii) diffraction at the entrance slit; and (iii) diffraction at the working apertures of the objectives and the prism. Aberrations of imaging systems are described in detail in Chapter 2. As to diffraction, we will summarize here, for the reader's convenience, some facts and formulas relating to diffraction phenomena. As is well known (e.g., see detailed explanation in Born and Wolf, 1968) diffraction of light on a rectangular aperture of width *b* causes light waves to propagate in different directions after the aperture, even if the incident radiation constitutes a parallel beam. Figure 5.13 demonstrates the phenomenon and the light intensity as a function of diffraction angle φ . This function is governed by the formulas:

$$I_{\varphi} = I_0 \frac{\sin^2(u)}{u^2};$$

$$u = \frac{\pi b \sin \varphi}{\lambda}$$
(5.14)

which has a maximum at $u = \varphi = 0$ and the first minima at $\sin(\varphi_1) = \pm \lambda/b$. Between two minima, $+\varphi_1$ and $-\varphi_1$, there is about 93% of the total energy transferred through the aperture.

If a circular aperture of diameter d is set in the parallel beam the diffraction pattern is similar to the one shown in Fig. 5.13, but its analytical description differs from Eq. (5.14): the function $\sin(u)$ is substituted by Bessel functions of the first order and the first minima appear at the angle $\sin(\varphi_1) = 1.22\lambda/d$. If a lens of focal length f' is positioned after the aperture it concentrates the diffracted beams in its focal plane, each ray of direction φ coming to the point with



FIGURE 5.13 Diffraction at a rectangular aperture: (a) propagation of light; (b) angular distribution of intensity.

radial coordinate $r = f' \tan(\varphi)$ so that the intensity distribution I(r) is an axially symmetrical function (the so-called Airy function mentioned in Chapter 2).

Returning to the spectral instrument shown in Fig. 5.11 and supposing the entrance slit of infinitesimal width and both objectives and the prism working with a beam of effective aperture D and being diffraction limited, one can expect that due to diffraction at D a spot of size $s'_{dif} = 2 \lambda / Df'_2$ is the minimum spot size in the output plane T achievable in the system (for a single wavelength). If the entrance slit has finite width s each part of s creates the same diffractive spot s'_{dif} centered in a slightly different coordinate, each one being the geometrical image of the corresponding point of the entrance slit. Altogether they create a spot of a finite width with an intensity distribution like that shown on the left-hand side of Fig. 5.14 (curve 1). The smaller the entrance slit width the narrower the output spot (curves 2 and 3). The maximum intensity in the spot center remains unchanged, until the geometrical size becomes equal to s'_{dif} (curve 3). A further reduction of the entrance slit is accompanied by a decrease of the spot intensity (curves 4 and 5). This can be explained by taking into account diffraction at the entrance slit itself. That is, if the cross-section of the diffracted beams in the plane of L_1 is larger than the effective size D of the objective, then part of the energy is lost. Therefore, with regard to energy transfer the best situation is achieved when $2\lambda/bf'_1 = D$. However, the highest spectral resolution requires a smaller size of spot. A compromise is achieved when the geometrical size of the image b' is equal to half of $s'_{\rm dif}$ – in this case about 83% of the energy transferred through the entrance slit participates in the creation of spectral lines and the maximum spectral resolution is still obtained (minimum $\delta\lambda$ according to the Rayleigh criterion). Such a case is presented in Fig. 5.15 and the corresponding value

$$b = \frac{\lambda}{D} f_1' \tag{5.15}$$

is usually chosen as the optimal size of the entrance slit.

Since the effective working diameter D is determined by the prism size and refractive angle, the resolution \Re of the prism-based device with diffraction-limited



FIGURE 5.14 Light intensity of the entrance slit image obtained in the output plane T.



FIGURE 5.15 Optimal size of entrance slit.

optics is determined, as can be shown, by the expression

$$\Re = B \frac{\mathrm{d}n}{\mathrm{d}\lambda} \tag{5.16}$$

where *B* is the base of the prism and $dn/d\lambda$ is the dispersion of the prism material.

In practice, very often aberrations of the optics as well as the final angular size of the light source illuminating the entrance slit, rather than diffraction phenomena, govern the spot size s'. In such a situation the intensity distribution across the exit spot is similar to curve 1 of Fig. 5.14. Then the minimum resolvable spectral interval and resolution of the system can be calculated from the real size of the spot and the reciprocal linear dispersion of the system (e.g., see Problem 5.15).

Problems

5.12 Find an analytical expression for the angular dispersion of a prism working around the minimum deviation angle.

5.13. The dispersion element of a spectrometer is a 40° prism made of BK-7 glass $(n_d = 1.5163, n_F - n_C = 0.008054)$ with a base of 30 mm. The instrument is intended for visible wavelengths. The optics of the device is diffraction limited and it comprises two identical lenses of 40 mm diameter and 200 mm focal length.

- (a) Find the resolution of the spectrometer and the minimum resolvable spectral interval.
- (b) Calculate the optimal size of the entrance slit.
- (c) If a stop of 30 mm diameter is positioned in front of L_1 how will it affect the resolution? What happens to the spectrum if the entrance slit



FIGURE 5.16 System for testing pollutants in water.

remains unchanged? What should be the optimal width of the slit if the stop is present?

5.14. The spectral system shown in Fig. 5.16 is exploited for revealing pollutants in water supplied through a transparent vessel M. The system is configured around a prism P made of BK-7 glass (see the glass data in Problem 5.13) and having a refraction angle $\alpha = 60^{\circ}$ and base B = 60 mm. Also included in the system are two identical lenses, L₁ and L₂, of focal length 300 mm (diffraction limited) and a line detector array T with pixels of 10 µm pitch. Illumination originates in a source S of a narrow angular size and it is projected onto an entrance slit b. The source has a high-intensity spectral line of $N_1 = 16,800$ cm⁻¹.

- (a) Calculate the resolution of the system for a given spectral line.
- (b) Is it possible to detect a pollutant having a typical spectral line of $N_2 = 16,790 \text{ cm}^{-1}$?
- (c) If the prism is replaced by another one of the same geometry but made of SF-5 glass ($n_d = 1.6727$; $n_F n_C = 0.020884$), how will this influence the answer to (b)?

5.15. The system shown in Fig. 5.16 works with a light source of a finite angular width and imaging optics with aberrations that cannot be neglected. As a result, the minimum spot in the output plane T is of 30 μ m. Calculate the system resolution for the wavenumber given in Problem 5.14 and check the answers to the last two questions of that problem.

5.3. Diffraction Gratings and Grating-based Systems

5.3.1. Plane Diffraction Gratings and Related Configurations

In most spectral instruments a diffraction grating is used as a dispersive element. To explain its features and operating mode we will start with the simplest case of a plane transparent grating made of N parallel narrow slits of precisely equal width b illuminated by a parallel monochromatic light beam. The slits are separated by non-transparent areas of size a, so that a periodic structure of spatial period d = b+a is established. An example of such a periodic transparent–non-transparent structure is the so-called Ronchi rolling plate, although usually these plates are less accurate than the real diffraction gratings used in spectral instruments.

A lens of focal length f' is positioned after the grating (see Fig. 5.17). Due to the diffraction phenomenon numerous diffracted beams are created after each slit and rays of the same direction are concentrated by the lens in a single point in the lens focal plane (separate points for each diffracted angle φ). A pair of two such rays propagating in the same direction and belonging to two adjacent slits have an optical path difference $AB = \Delta = d \sin \varphi$, as shown in Fig. 5.17a. N pairs coming to the same point in the focal plane are coherent with each other and therefore a multi-beam interference pattern is created here. As a result, the light intensity in the lens focal plane is described by the function

$$I(x) = I_0 f' \frac{\sin^2 u}{u^2} \cdot \frac{\sin^2(Nv)}{(\sin v)^2}$$
(5.17)



FIGURE 5.17 (a) Diffraction of a plane diffraction grating and (b) intensity distribution of diffracted light.

where *u* is the same as in Eq. (5.14) for a single slit of width *b*, $u = \pi b \sin(\varphi)/\lambda$, and $v = \pi d \sin(\varphi)/\lambda$ is determined by the grating period, *d*. Figure 5.17b demonstrates the intensity distribution as a function of diffraction angle φ . The following features are of importance:

(a) The second term on the right-hand side of Eq. (5.17) reaches maximum values when $\sin v = 0$ and $\sin(Nv) = 0$ simultaneously, which correspond to discrete directions $\varphi_{\text{max}}^{(m)}$ obeying the following conditions:

$$\sin(\varphi_{\max}^{(m)}) = m\lambda/d \quad (m = 0; \pm 1; \pm 2; ...).$$
 (5.18)

These maxima are called the principal maxima of diffraction order m (m = 0) yields the "zero-order maximum" corresponding to $\varphi_{\text{max}}^{(0)} = 0$; m = 1 and m = -1 yield the "first-order maximum" and the "minus first-order maximum" corresponding to directions $\sin(\varphi_{\text{max}}^{(1)}) = \lambda/d$ and $\sin(\varphi_{\text{max}}^{(-1)}) = -\lambda/d$; etc.).

(b) The width of each principal maximum is determined by the two closest minima, one on the left and the other on the right of the maximum, each minimum being related to the increment δΔ = ±λ/2 in the optical path difference between two rays 1 and 2 (increment of ±π in the argument of sin(*Nv*)). In terms of diffraction angle this gives:

$$\delta \sin(\varphi) = \pm \lambda / (Nd). \tag{5.19}$$

(c) The first term on the right-hand side of Eq. (5.17) is related to diffraction at a single slit of width *b* and it determines the envelope (dotted line in Fig. 5.17b) of the light intensity of the principal maxima. The minimum of the envelope corresponds to the directions

$$\sin(\varphi_{\min}) = \pm \lambda/b \tag{5.20}$$

and therefore all significant principal maxima are concentrated inside the cone of diffraction angles given by Eq. (5.20).

If the grating is illuminated by non-monochromatic light each wavelength creates a separate diffraction pattern. Conditions (5.18)–(5.20) remain valid, and therefore the zero-order direction (m = 0) is common for all wavelengths (all of them come to the focus of the lens) whereas the location of the principal maxima in any other diffraction order ($m \neq 0$) depend on λ . Hence the wavelengths are separated in the focal plane of the lens and monochromatic light can be obtained by using, for instance, a narrow slit moving in this plane from one position to another in the vicinity of principal maximum of order m. The higher the diffraction order exploited, the greater the separation of a pair of wavelengths: $m\lambda_1/d - m\lambda_2/d$.



FIGURE 5.18 (a) Diffraction of a reflective plane grating and (b) intensity distribution of diffracted light with energy concentration in the second order.

Very close wavelengths are overlapped and the resolution ability of the grating is again defined according to the Rayleigh criterion (see Fig. 5.12).

In reality reflection diffraction gratings are used instead of transparent ones. Such a grating is shown in Fig. 5.18a. In this case *N* identical parallel grooves, instead of *N* parallel slits, constitute the grating. Each groove is characterized by three parameters: *d*, the total width of the groove, *b*, the width of a single reflective element (a mirror); and γ , the inclination angle of each small mirror (sometimes called the "blazing angle"). It is these parameters which enable one to improve significantly the efficiency of the diffraction grating, as explained below.

To understand the operation and advantages of a reflection grating we will address again the optical path difference between two parallel rays, 1, and 2, of two adjacent grooves:

$$\Delta_{21} = AD - BC = d(\sin\varphi - \sin\psi)$$

for any chosen incident angle ψ and diffraction angle φ . In Fig. 5.18 $\psi > 0$ and $\varphi > 0$, but, in general, one should keep in mind that while doing calculations for reflective gratings the sign conventions for all considered angles have to be applied with care (see examples and explanation in Problem 5.16).

Expression (5.17) is also valid for a reflective grating, although the directions of the principal maxima, instead of Eq. (5.18), are governed by the formula

$$\sin(\varphi_{\max}^{(m)}) = \sin(\psi) \pm m\lambda/d.$$
(5.21)

The width of these maxima still obey Eq. (5.19). Instead of Eq. (5.20) we have the following expression describing the diffraction minimum of a single mirror:

$$b(\sin\beta - \sin\alpha) = \pm\lambda. \tag{5.22}$$

The low efficiency of a transparent grating for spectral measurements results from the fact that the maximum of energy is concentrated in the zero order (see Fig. 5.17b) which is useless for spectral resolution. A reflective grating enables one to optimize the distribution of energy between the diffraction orders. That is, it can be designed in such a manner that maximum energy will go into the diffractive order chosen for normal operation of an instrument and the zero order is minimized. To do this one can optimize the parameters of the grating grooves. Taking into account that maximum energy is concentrated around the direction of specular reflection of each small mirror which corresponds to the condition $\alpha = -\beta$ and also obtaining from Fig. 5.18a

$$\psi = \alpha + \gamma; \quad \varphi + \gamma = \beta \tag{5.23}$$

we use both conditions in Eq. (5.21) to obtain for a chosen value *m*:

$$2\sin(-\gamma)\cos(\psi - \gamma) = m\lambda/d.$$
(5.24)

This equation allows one to calculate the optimal angle of the grooves inclination, γ .

Furthermore, substituting α and β from Eq. (5.23) in Eq. (5.22) and keeping in mind that for the zero-order direction $\varphi_{\text{max}}^{(0)} = \psi$, we get

$$2\cos\psi\cdot\sin\gamma = \lambda/b. \tag{5.25}$$

Equation (5.25) allows one to calculate the active size of each mirror of the grooves, *b*. If γ and *b* obey Eqs. (5.24) and (5.25) the maximum energy in reflected light is concentrated in a chosen diffractive order *m* and the zero order is minimized. Such an example is presented in Fig. 5.18b for the case of m = -2.

An architecture of a spectral instrument where a plane reflective grating is used as a dispersive element is shown in Fig. 5.19. Radiation of a light source A is directed by a mirror M towards an entrance slit S. Since S and the output detector array, T, are positioned in the focal planes of lenses L_1 and L_2 , the grating G is obviously operating with parallel beams.

The main characteristics of the device can be calculated in the following manner. The angular dispersion is found by differentiation of the condition of the principal maxima (Eq. (5.21)) at any given $\psi = \text{const}$:

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = \frac{m}{d\cos\varphi}.\tag{5.26}$$



FIGURE 5.19 Basic configuration of a spectrometer with a plane reflective grating.

Hence, for the linear dispersion we have

$$\frac{dl}{d\lambda} = \frac{m}{d\cos\varphi} f_2'. \tag{5.27}$$

Using Eq. (5.17) for the width of a principal maxima and taking into account, as usual, the Rayleigh criterion, one can find the minimum resolvable spectral interval:

$$\delta \lambda = \frac{(\lambda/Nd)f_2'}{dl/d\lambda} = \frac{\lambda \cos \varphi}{mN}$$
(5.28)

which gives for the resolution

$$\Re = \frac{\lambda}{\delta\lambda} = \frac{mN}{\cos\varphi}.$$
(5.29)

This is actually a theoretical limit for an aberration-free system and with the whole grating participating in producing the diffraction pattern.

The optimal width of the entrance slit can be found in exactly the same way as in the case of a prism-based configuration (see Section 5.2, Eq. (5.15)), providing the effective diameter of the lenses, *D*, is compatible with the total width of the grating:

$$D = B\cos\psi = Nd\cos\psi \tag{5.30}$$

and the imaging optics is diffraction limited (aberrations can be neglected).

If optical aberrations and illumination conditions are taken into account Eqs. (5.28) and (5.29) are not useful and the resolution and limiting spectral interval should be based on the actual size of the spot in the output plane, as explained in Section 5.2.

A cost-effective configuration for a grating-based instrument is demonstrated in Fig. 5.20. Such an architecture, known as an autocollimating scheme, exploits a single lens L for both illumination of the reflective grating G and production of the spectrum in the output plane T where either a detector array or an exit



FIGURE 5.20 Architecture of an autocollimation spectrometer with reflective grating.

slit is positioned. A small prism (or mirror) P tilts the incident beam coming from the entrance slit S towards the lens L and the grating G.

Problems

5.16. Show that the optical path difference (OPD) between two parallel rays 1 and 2 incident on a reflective diffraction grating (see Fig. 5.18a) obeys the expression $\Delta_{21} = d(\sin \varphi - \sin \psi)$ for any incident angle ψ and diffraction angle φ . [Note: Consider positive and negative angles using the sign convention described in Chapter 1.]

5.17. Find the relation between the working order of diffraction, *m*, and the total spectral interval, $\Delta \lambda = \lambda_{\text{max}} - \lambda_{\text{min}}$, of a spectrometer or a monochromator allowing one to avoid overlapping between adjacent diffraction orders.

5.18 Optimization of reflective grating. Find the optimal parameters of a reflective grating of 25 mm total size working in the second diffraction order in visible wavelengths and providing a minimum resolvable spectral interval of 0.2 Å. The grating is illuminated by a parallel beam incident at an angle $\psi = -15^{\circ}$.

5.19. Spectrometer with autocollimation architecture. A schematic configuration of an instrument is presented in Fig. 5.20. Assuming that the grating G of 300 lp/mm and 1" size is tilted at 15° to the optical axis and located 100 mm from lens L (diameter D = 100 mm, focal length of 1,200 mm), and supposing the system is intended for operation in the red and near-infrared wavelengths (from 600 nm up), find:

- (a) the maximum achievable (theoretical) spectral resolution and the free spectral range without overlapping;
- (b) the location of the detector array, the size of a single pixel, and the maximum useful number of pixels.

5.20. A monochromator designed for visible wavelengths comprises a reflective diffraction grating of 1,200 lp/mm illuminated at 10° and a camera lens

of 150 mm focal length. The grating is optimized for the first diffraction order. Assuming the exit slit is 20 μ m in width, find the spectral interval of radiation emerging from the device when it is set for $\lambda = 500$ nm.

5.21. A spectrometer with a 1" reflective grating of 600 lp/mm is chosen as a tool for investigating the stability of a diode laser of 0.83 μ m working wavelength. The laser beam is collimated, so that the spot incident on the grating is 2 mm in size. Find the minimum achievable spectral width of the line at the exit of the spectrometer. Does the instrument suit the purpose of the study?

5.3.2. Systems with Concave Diffraction Gratings

A concave diffraction grating is actually a combination of the reflective grating described in Section 5.3.1 and a concave mirror. Spectral devices with mirror imaging optics instead of lenses are usually exploited if spectral measurements are to be done at UV or IR wavelengths where absorption of the lens material might affect significantly the propagation of radiation. Use of a concave grating allows for further simplification of the system since it reduces the overall number of elements.

One of the simplest configurations is presented in Fig. 5.21. The image of an entrance slit S is created by a grating G in location S' where an exit slit or a detector array can be positioned. The grating is imposed on the surface of a spherical mirror of curvature ρ . It can be shown that both S and S' are located on the same circle (the so-called Rowland circle) of diameter ρ and the distances *r* and *r'* from S to G and from G to S', respectively, are calculated for a chosen incident angle ψ and diffraction angle φ as follows:

$$r = \rho \cos(\psi); \quad r' = \rho \cos(\varphi).$$
 (5.31)

All relations governing the behavior of a plane diffraction grating also remain valid for a concave grating as well: the angular position of the principal maxima and



FIGURE 5.21 Configuration of a monochromator with a concave diffraction grating.

their widths are defined by Eqs. (5.21) and (5.22); the optimal parameters of the grating grooves are calculated from Eqs. (5.24) and (5.25); and angular dispersion and resolution are as in Eqs. (5.26), (5.28), and (5.29). As to the linear dispersion of a concave grating, we have to take into account the linear magnification, V, existing in the configuration shown in Fig. 5. 21:

$$V = \frac{r'}{r} \tag{5.32}$$

and therefore

$$\frac{\mathrm{d}l}{\mathrm{d}\lambda} = \frac{mr'}{d\cos\varphi} = \frac{m\rho}{d}.$$
(5.33)

If a simple configuration like that of Fig. 5.21 is used in the design of a UV monochromator it should be taken into account that variation of wavelength will require not only movement of the grating (it should be rotated around the vertical axis parallel to the grooves), but also will require displacement of the exit slit – this should remain on the Rowland circle which is moved together with the grating. This leads to difficulties in the mechanics of the system and is one reason why other configurations have become more popular. One of them is shown in Fig. 5.22. Here a slit S is located in the focus of a spherical mirror M which generates a tilted parallel beam incident on a concave grating G. The spectrum is created on a cylindrical surface of radius

$$r' = \frac{\rho \cos^2(\varphi)}{\cos \psi + \cos \varphi}.$$
(5.34)

This value should be taken into account while calculating the optical magnification of the system and the optimal width of the entrance slit.

Problems

5.22. A monochromator for UV wavelengths (2,000–4,000 Å) includes a concave grating of 2,400 lp/mm made on a mirror surface of 50 cm curvature and optimized



FIGURE 5.22 Concave grating in a parallel beam.

for diffraction order m = -1. The grating is illuminated by a beam incident from a direction $\psi = 30^{\circ}$. Find the positions of the entrance and the exit slits with regard to the grating and calculate the width of the exit slit if the entrance slit is 0.2 mm in size.

5.23. A UV spectrometer has a concave grating of 600 lp/mm grooves and 250 mm curvature. The entrance slit is 125 mm aside from the center of the grating. At the output plane a CCD line detector of 1,024 pixels, 15 μ m each, is positioned. The spectrometer is aligned in such a way that a wavelength of 300 nm in the diffraction order m = -2 is concentrated at the center of the detector array. Find:

- (a) the location of the CCD array relative to the grating;
- (b) the theoretical spectral resolution and the total spectral range covered by the device.

[Note: Theoretical values are calculated assuming that the imaging is aberration free and the grating is ideal.]

5.4. Interferometry-based Spectral Instruments

5.4.1. Interference Filters and Fabry–Perot Interferometer

Spectral instruments designed according to interferometry architecture constitute the group of super high-resolution systems. We start with a simple interference filter and then describe a system with a Fabry–Perot etalon.

An interference filter is usually a flat slab of glass coated on both sides with highly reflective coatings with very low absorption. Due to the high reflectivity on both sides any incident beam passing through the coating is multiply reflected inside the slab, each reflection being accompanied by the generation of a new ray going out of the slab, as shown in Fig. 5.23. The optical path difference between two adjacent rays, say 1' and 2', is expressed as

$$\Delta = AB + BC - AD = 2tn\cos(r) \tag{5.35}$$

where t is the thickness of the slab, n is its refractive index, and r is the angle of refraction inside the slab (related to the incident angle, i, as usual, by Eq. (1.2)).

Let the slab be illuminated by a parallel monochromatic beam of wavelength λ . Since all rays emerging from the slab originate from the same incident ray 1 (i.e., from the same incident wave front) they all are coherent with each other and therefore when gathered by a lens they will interfere. Denoting the coating transmittance and reflectivity as τ and R, respectively, one obtains the following intensity distribution of the resulting interference pattern (see Fig. 5.23b; for



FIGURE 5.23 (a) Generation of multiple rays in an interference filter and (b) intensity distribution of transmitted light.

details, see Born and Wolf, 1968):

$$I(r;\lambda) = I_0 \frac{\tau^2}{(1-R)^2 + 4R\sin^2(\Phi/2)}$$

$$\Phi = \frac{2\pi}{\lambda} \Delta = \frac{4\pi}{\lambda} tn \cos(r).$$
(5.36)

Given the thickness and refractive index of the slab, the intensity variation as a function of wavelength (if the incident angle is constant) or as a function of incident angle i (and r) if the wavelength is not changed can be calculated. In Fig. 5.23b the graph of intensity according to Eq. (5.36) is presented. The maximum and minimum values of Eq. (5.36) are

$$I_{\max} = \frac{\tau^2}{(1-R)^2}; \quad I_{\min} = \frac{\tau^2}{(1+R)^2}.$$
 (5.37)

Evidently there are a number of maxima, obeying the condition $\Delta = m\lambda$ (m = 1; 2; 3...). What is important for spectral measurements is the width of the graph around the maximum. This value, $\delta\lambda$, is usually defined as the segment where $I \ge 0.5I_{\text{max}}$ and is called the bandpass or FWHM (full width at half maximum). For a chosen angle of incident radiation (*i*) this means that the spectral interval, $\delta\lambda$, is described as

$$\delta \lambda = \frac{\lambda}{mN_{\rm e}} \tag{5.38}$$

where

$$m = \Delta/\lambda = \frac{2tn\cos r}{\lambda}; \quad N_{\rm e} = \frac{\pi\sqrt{R}}{1-R}.$$
 (5.39)

The value N_e is called the effective number of rays participating in interference and *m* is the interference order. The higher the coating reflection *R*, the greater the number of relevant rays N_e and the smaller the spectral interval $\delta\lambda$ transmitted by the filter.

To achieve better spectral resolution it is possible to put several interference filters in sequence, one after another. Each ray emerging from the first filter (called sometimes a first cavity) generates in the second filter (second cavity) a multi-reflection pattern like that shown in Fig. 5.23a. The total transmitted intensity distribution can be described, approximately, as Eq. (5.36) in power k, where k is the number of cavities in the sequence. Hence, the greater the value of k the smaller the value of $\delta\lambda$ that can be achieved. In practice all cavities are arranged as a multi-layer coating on a single substrate and such an arrangement is called a multi-cavity interference filter.

A Fabry–Perot etalon is actually two mirrors of very high reflectivity precisely parallel to each other and separated either by air or by a transparent solid (usually glass or quartz). If radiation is transmitted through the etalon a multiple-ray interference pattern is created, just as in the case of the interference filter described above. Specific features of the etalon are exploited in order to establish a spectral system of extremely high resolution. The configuration of such a system is presented in Fig. 5.24a. Two lenses, L_1 and L_2 , build the image of the entrance aperture S in the exit plane with aperture S'. The etalon is positioned in the parallel beams propagating between the two lenses.

The inner sides of the etalon plates are precisely aligned. The outer surfaces are tilted to the optical axis in order to avoid the influence of undesirable reflections from these surfaces. The tested light source A and its optics M illuminate the entrance aperture with radiation of some kind of monochromaticity (see below). Since the mirror separation, t, is much greater than the wavelength, the order of interference, m, is very high. It is evident that m_{max} corresponds to the beam parallel



FIGURE 5.24 (a) Configuration of a spectral system with a Fabry–Perot etalon and (b) the interference pattern at the exit aperture S'.

to the optical axis (r = 0) and originates from the central point of aperture S. The next maximum of the same wavelength λ is obtained at $m_1 = m_{\text{max}} - 1$ corresponding to the beam tilted at some angle r_1 and originating from a point displaced from the optical axis. All points displaced symmetrically constitute in the output aperture S' a ring of light corresponding to the order m_1 . This procedure is valid also for all other interference orders and the whole picture looks like that in Fig. 5.24b. The radius of the *k*-th ring, ρ_k , can be calculated as follows (we assume here n = 1):

$$\rho_k = f' \sqrt{\frac{2k}{m}} = f' \sqrt{\frac{k\lambda}{t}}.$$
(5.40)

As we see, the distance between adjacent rings decreases on moving away from the axis.

The minimum resolvable spectral interval, $\delta\lambda$, is defined by Eq. (5.38). Thus, for the resolution of the spectral instrument we have

$$\Re = \frac{\lambda}{\delta\lambda} = mN_{\rm e}.\tag{5.41}$$

It is evident from Eq. (5.40) that the location of interference rings depends on λ and in order to avoid overlapping of interference patterns belonging to adjacent interference orders the total spectral interval, $\Delta\lambda$, of the light source participating in the creation of the rings should obey the following condition:

$$\Delta \lambda = \frac{\lambda^2}{2t}.$$
(5.42)

Problems

5.24. An interference filter is built around a transparent dielectric layer of 0.2 μ m width and refractive index n = 1.4. An optical coating provides equal reflectivity on both sides of R = 0.95. Find the wavelength of maximum transmittance and the FWHM of the filter.

5.25. How much will the maximum transmittance and FWHM be affected if the interference filter of Problem 5.24 is tilted at 20° to the incident radiation? Or even 30° ?

5.26. Is it possible to get an interference filter with 0.6 μ m working wavelength and 1 nm bandpass?

5.27. An interference filter designed for a wavelength of 0.5 μ m and bandpass FWHM of 5 nm is illuminated by a convergent beam. The convergent angle is 20°. What is the effective bandpass for transmitted radiation in such a case?

5.28. An investigation of physical processes in an electric discharge arc is based on the detection, with a Fabry–Perot etalon, of the shape of the two closest spectral lines of the iron triplet: $v_1 = 32,258 \text{ cm}^{-1}$ and $v_2 = 32,255 \text{ cm}^{-1}$. What should be the air spacing between the mirrors of the instrument and what is the spectral resolution achievable in the system?

5.29. The spectral analysis system shown in Fig. 5.24 comprises a Fabry–Perot etalon with glass spacing of 1.7 mm, two lenses of 200 mm focal length, and two apertures, S and S', of 7 mm diameter each. The etalon mirror coating provides a reflection of 95% in the visible region. Calculate:

- (a) the order of interference and the spectral resolution for $\lambda = 0.5 \ \mu m$;
- (b) the radius and the width of the first and second interference rings in the output plane;
- (c) if there is any wavelength in the visible for which two rings can be observed in the aperture S'.

5.4.2. Fourier Spectrometer

190

A Fourier spectrometer combines the advantages of a highly sensitive interferometric system with an advanced signal processing technique. Basically it is a dual-beam interferometer with variable optical path difference, a single light detector, and digital electronic circuitry for fast Fourier transform computations. In describing the main principle we consider the Michelson interferometer architecture shown schematically in Fig. 5.25. Radiation of the tested light source, A, is concentrated by an illumination lens L_1 on an entrance stop S which is located in the focal plane of lens L_2 . After lens L_3 , also in its focal plane, an exit stop S' is positioned followed by a detector D. Between the two lenses parallel light beams are propagated. The beam coming from L_2 is split by a beam splitter BS, one part going to mirror M_1 and the other to mirror M_2 . After reflection by the mirrors, both parts meet each other in the plane of BS and proceed further to lens L_3 and to detector D. Interference takes place in the plane of BS and thereafter, and its result depends on the optical path difference between the two branches of the interferometer.

Initially the distance from BS to M_2 is equal exactly to the distance from BS to M_1 , and there is no phase difference between two monochromatic beams (of wavelength λ) coming to D. The light intensity, I_{λ} , and the detector signal, i_D , will achieve the maximum value at this moment. Then mirror M_2 starts moving along the horizontal axis at a constant speed V. This results in a change of the optical path difference between the two branches and therefore the light intensity and the detector signal will vary accordingly. Since radiation passes each additional



FIGURE 5.25 Configuration of a Fourier spectrometer.

segment twice, moving the mirror by a quarter of a wavelength will reduce the interference intensity to a minimum and further moving an additional $\lambda/4$ will increase the intensity again to the maximum, and so on. Hence, the optical path difference, Δ , varies in time as $\Delta(t) = 2Vt$ and the interference intensity at the exit stop is

$$I_{\lambda}(t) = 2I_{0\lambda} \left[1 + \cos\left(\frac{4\pi}{\lambda} V t\right) \right]$$

and therefore the variable (AC) detector signal will be

$$i_{\rm D}(t) = 2\Re I_{0\lambda} \cos(\omega_0 t) \tag{5.43}$$

where \Re is the detector responsivity (see Section 4.1), $\omega_0 = 4\pi V N_0$, and N_0 is the wavenumber of propagated monochromatic radiation (dimensions of V are cm/s and N_0 are cm⁻¹).

Let M₂ move between two extreme positions, P₁ and P₂, with coordinates $x_1 = a - Vl/2$ and $x_2 = a + Vl/2$ during the time interval, *T*:

$$T = l/V. \tag{5.44}$$

Therefore $i_D(t)$ is a finite function obeying Eq. (5.43) in the time interval [0;*T*] and it can be prolonged by a zero value outside of this interval. It is well known

that the Fourier transform of such a function is expressed as follows:

$$F(\omega) = 2\Re I_{0\lambda} \int_{-\infty}^{\infty} \cos(\omega_0 t) \exp(-j\omega t) dt$$
$$= \operatorname{const} \left\{ \frac{\sin[(\omega_0 - \omega)T/2]}{\omega_0 - \omega} + \frac{\sin[(\omega_0 + \omega)T/2]}{\omega_0 + \omega} \right\}.$$
(5.45)

The graphical representation of Eq. (5.45), called the Fourier transform spectrum, for positive frequencies is a single spectral line of the shape shown in Fig. 5.26a. It corresponds to a single wavenumber of monochromatic radiation. If a number of wavelengths are emitted simultaneously by the light source A each one generates a separate interference pattern, but all of them arriving simultaneously at the detector cause the complex signal

$$i_{\rm D}(t) = 2 \sum_{n} \Re_n I_{0n} \cos(4\pi \nu_n V t)$$
 (5.46)

and the corresponding Fourier transform is

$$F(\omega) = \sum_{n} C_n \frac{\sin[(\omega_n - \omega)T/2]}{\omega_n - \omega}.$$
(5.47)

The case shown in the Fig. 5.26b demonstrates the spectrum of the propagated beams and relates to all n wavelengths presented there.

Calculation of the Fourier transform is a cumbersome and time-consuming procedure, but since the advent of the FFT (fast Fourier transform) algorithm this operation is easily done using digital electronic circuitry which processes the output detector signal.



FIGURE 5.26 (a) Fourier spectrum of single-wavelength and (b) multiple-wavelength radiation.

The resolution of the system depends on the width of a single spectral line as described by Eq. (5.45). The half width of the line shape is defined by the first minimum for which we have $(\omega_0 - \omega)T/2 = \pi$ and therefore

$$N_0 - N = \frac{1}{2VT} = \frac{1}{2l}.$$
(5.48)

This means that the minimum resolvable spectral interval depends on the range of movement of mirror M_2 .

Problems

5.30. A Fourier spectrometer is operated in the near-IR wavelength range from 1 to 5 μ m and provides a spectral resolution (minimum resolvable spectral interval) of $\delta\lambda = 0.2$ nm for all wavelengths. Find the spectral resolution in wavenumbers for minimum and maximum wavelengths; and the maximum optical path difference (OPD) and the range of the scanning mirror movement.

5.31. The Fourier spectrometer mentioned in Problem 5.30 exploits an InSb detector with electronic circuitry of 20 kHz bandwidth. Assuming that the spectral resolution is 0.2 nm for a wavelength of 5 μ m, find the scanning speed required for normal operation of the apparatus.

5.5. Spectrophotometry

Spectrophotometers are devices for investigating the transmission or reflection of samples of materials at different wavelengths, primarily with the aim of measuring the concentration of some components in complex mixtures of liquids, gases, or solids. They are widely used in biological and medical applications as well as in chemical technology and industrial laboratory testing.

As mentioned in Section 5.1, the intensity of radiation propagated through a slab of material characterized by an absorption factor α is reduced exponentially according to Bouguer's law (Eq. (5.10)). If the absorption centers are spread in a transparent medium (like a dilute solution) and the volume concentration of absorbing particles in the medium is *C* then the absorption factor of the medium, α_M , is governed by Beer's law:

$$\alpha_M = \alpha C. \tag{5.49}$$

This rule of linear proportion between absorption of the medium and concentration of absorbing species has been examined in many studies and verified for a wide



FIGURE 5.27 Configuration of a two-channel spectrophotometer.

range of concentrations. Thus, an unknown concentration of absorbing particles can be found from Eq. (5.49) if $\alpha_{\rm M}$ is measured and α of a single absorption center (a particle or collection of identical molecules) is known in advance.

Numerous configurations can be used for spectrophotometric measurements. An example of a two-channel spectrophotometer is presented in Fig. 5.27. Monochromatic light originating in a system monochromator M is split into two beams by a beam splitter and then, after transmission through two samples of a test material, S_1 and S_2 , is focused on detectors D_1 and D_2 . Let the thickness of the samples be t_1 and t_2 . Then the intensity of light coming to the detector in each channel is described as

$$I_1 = I_{01}(1-R)^2 \exp(-\alpha_M t_1)$$
$$I_2 = I_{02}(1-R)^2 \exp(-\alpha_M t_2)$$

where *R* is the reflection on each side of the sample. Assuming that the radiation is divided equally between two channels ($I_{01} = I_{02}$) and all optical elements and the detectors in both channels are also identical, we derive from the last two expressions:

$$I_1/I_2 = T_1/T_2 = T_{21} = \exp[-\alpha_{\rm M}(t_1 - t_2)]$$
(5.50)

which gives

$$\alpha_{\rm M} = \frac{1}{t_2 - t_1} \ln(T_{21}). \tag{5.51}$$

To use the measured value α_M in Eq. (5.49) we should also keep in mind that α is related to the optical constants of the substance as in Eq. (5.11). In the

case that the optical constants for the absorbing centers are not known at least one calibration experiment has to be carried out prior to using the instrument for routine measurements. In this calibration experiment the concentration of the absorbing particles should be known and be identical for both channels. This enables one to measure α_M and then to find from Eq. (5.49) the value of α which can be used in further measurements.

Problems

5.32. A two-channel spectrophotometer, like that of Fig. 5.27, was used for measurement of the concentration of Ag particles homogeneously dispersed in a partially transparent solution. Two cuvettes filled with the solution were introduced in the device. The thickness of the liquid was $t_1 = 1.0$ mm in the first vessel and $t_2 = 3.5$ mm in the second. The ratio of the detector signals measured in both channels for a wavelength of 0.59 µm was $i_{\text{Det}2}/i_{\text{Det}1} = 0.05$. Calculate the concentration of Ag in the solution.

[Note: Refractive index of Ag particles in the spectral interval of the measurements is n = 0.18 - 3.64j.]

5.6. Solutions to Problems

5.1. According to the definition of wavenumber, we obtain for the given spectral line $N = 10,000/0.546075 = 18,312.50 \text{ cm}^{-1}$ and the optical frequency

$$\nu = \frac{c}{\lambda} = \frac{3 \times 10^{10} \text{ cm/s}}{0.546075 \times 10^{-4} \text{ cm}} = 5.493751 \times 10^{14} \text{ Hz}.$$

Taking into account Plank's constant $h = 6.625 \times 10^{-34}$ J s and the conversion factor between J and eV (1 eV = 1.6022×10^{-19} J) we obtain the energy of the transition as $\Delta E = h\nu = 2.2716$ eV.

5.2. The natural width of the spectral line is 1.27×10^{-4} Å (see Section 5.1). From the definition of wavenumber one obtains $\Delta N/N = \Delta \lambda/\lambda$ and therefore

$$\Delta N = \frac{10,000}{\lambda^2} \Delta \lambda = \frac{1.27 \times 10^{-4}}{0.36} = 3.5278 \times 10^{-4} \,\mathrm{cm}^{-1}.$$

Since $\nu = c/\lambda$ we also find

$$\Delta \nu = \frac{c}{\lambda^2} \Delta \lambda = \frac{3 \times 10^{10} \text{ cm/s}}{0.36 \times 10^{-8} \text{ cm}^2} \times 1.27 \times 10^{-12} \text{ cm} = 10.58 \text{ MHz}.$$

5.3. Expression (5.3) yields for iron (atomic weight 56) at temperature 10,000 K:

$$\delta\lambda_{\rm D} = 7.18 \times 10^{-7} \times 3,100 \times \sqrt{\frac{10,000}{56}} = 0.0297 \text{ Å}$$

This value is significantly smaller (more than twice) than the wavelength differences of the triplet: $(\Delta \lambda)_{12} = 0.36$ Å; $(\Delta \lambda)_{23} = 0.34$ Å. Therefore, the triplet is still resolvable (if an appropriate spectral instrument is exploited).

5.4. Calculation of the Doppler broadening due to scattering on the electrons in the corona is done according to the relation presented in the problem (electron mass $m_e = 9.11 \times 10^{-31}$ kg; $k = 1.3806 \times 10^{-23}$ JK):

$$\Delta \nu_{\rm D}/\nu = \frac{1}{3 \times 10^8} \sqrt{\frac{2 \times 1.3806 \times 10^{-23}}{9.11 \times 10^{-31}}} \sqrt{600,000} = 0.014175.$$

Since $\Delta\lambda_D/\lambda = \Delta\nu_D/\nu$, we obtain $\Delta\lambda_D = 55.76$ Å which is about five orders of magnitude greater than the normal ("natural") width of the spectral line. Therefore absorption of photons occurs in wide spectral interval and this fact definitely can explain why Fraunhoffer absorption lines in the corona are so weak that they are hardly detectable.

5.5 Using Eq. (5.3) for Ne atoms (atomic weight 20) and remembering that the main line of a He–Ne laser is 6,328 Å, we get

$$\Delta \lambda_{\rm D} = 7.18 \times 10^{-7} \times 6328 \sqrt{\frac{350}{20}} = 0.019 \text{ Å}$$

which is about 200 times greater than the natural width of the spectral line.

5.6. The reflectance of each surface can be found from Eq. (5.9a) which yields the following: for Au, R = 84.9%; for Ag, R = 94.5%; for Cu, R = 73.2%; and for Ni, R = 61.9%.

5.7. The shortest wavelength corresponds to the greatest wavenumber, hence, using the definition of wavenumber, we find the reference wavelength as $\lambda_1 = 10,000/3,067 = 3.2605 \,\mu\text{m}$. The other wavelengths are $\lambda_2 = 3.2744 \,\mu\text{m}$; $\lambda_3 = 3.2982 \,\mu\text{m}$; $\lambda_4 = 3.3546 \,\mu\text{m}$; $\lambda_5 = 3.4247 \,\mu\text{m}$; $\lambda_6 = 3.4843 \,\mu\text{m}$. Denoting the coordinate of each wavelength λ_i in the output plane as x_i , one can calculate them with regard to the shortest wavelength as follows: $\Delta x_i = x_i - x_1 = (\lambda_i - \lambda_1)/(d\lambda/dl)$, where $d\lambda/dl = 50 \,\text{nm/mm}$. This gives $\Delta x_2 = 0.278 \,\text{mm}$; $\Delta x_3 = 0.754 \,\text{mm}$; $\Delta x_4 = 1.882 \,\text{mm}$; $\Delta x_5 = 3.284 \,\text{mm}$; $\Delta x_6 = 4.476 \,\text{mm}$.

5.8. Assuming that the wavelength difference between the two lines of the violet doublet represents the minimum resolvable spectral interval of the system,

 $\delta\lambda = 40$ Å, and calculating the resolution for wavelength of 4,022 Å, we get $\Re = \lambda/\delta\lambda = 101.5$.

To register properly two adjacent spectral lines incident on the CCD detector array we should require that the distance Δl between the wavelength centers will be equal to twice the pixel size at least – this requirement corresponds to the spatial sampling rate obeying the Nyquist theorem. In our case $\Delta l = 20 \,\mu\text{m}$ and therefore $dl/d\lambda = \Delta l/\delta\lambda = 0.5 \,\text{mm/nm}$ is the required linear dispersion of the system.

5.9. Each X-ray photon is of 20,000 eV energy in this case. Calculating the energy of generated light photons in eV gives:

$$\Delta E = \frac{hc}{\lambda} = \frac{6.625 \times 10^{-34} \times 3 \times 10^8}{0.39 \times 10^{-6} \times 1.6 \times 10^{-19}} = 3.185 \text{ eV}$$

and proceeding using the definition of the quantum efficiency, we get

$$\eta = 0.025 \frac{20,000}{3.185} = 157$$
 electrons/X-ray photon.

5.10. The required spatial resolution of 50 lp/mm dictates that the laser spot size will be as small as 50 μ m. Therefore, the total number of spots and the read-out sequences is $(250/0.05)^2 = 25 \times 10^6$. Since the total read-out time is no greater than 5 min, this means that a single spot read-out process must be finished after $\tau = 12 \,\mu$ s. Assuming that up to 90% of the F-centers of a single spot are read out $(N/N_0 = 0.1 \text{ at the end of the read-out process})$, we get $\ln(0.1) = -\sigma I \tau$ which results in the following:

$$I = \frac{2.303}{10^{-16} \times 12 \times 10^{-6}} = 1.919 \times 10^{21} \text{ photons/s/cm}^2 = 602.6 \text{ W/cm}^2$$

where we take into account that a single photon of the laser wavelength of 6,328 Å possesses energy of 3. 14×10^{-19} J. Keeping in mind that a single spot area is 1.963×10^{-5} cm² and only 75% of the laser photons achieve the PSL plate, we finally get the required power of the laser: $P = 602.6 \times 1.963 \times 10^{-5}/0.75 = 15.7$ mW.

5.11. Two spectral lines overlapping one another and obeying the Rayleigh criterion are shown in Fig. 5.28. Horizontal line coordinate, v, describes the location, x, in the output plane of the spectrometer (v is proportional to x and the proportionality factor depends on the focal length of the output lens of the system). We choose the origin of the variable v in the center of the first wavelength and the center of the second one is at the point v_0 . Then the total intensity, I_t , at each point is

$$I_{t} = I_{1}(v) + I_{2}(v) = I_{01} \frac{\sin^{2} v}{v^{2}} + I_{02} \frac{\sin^{2} (v - v_{0})}{v - v_{0}}$$



FIGURE 5.28 Problem 5.11 – Spectral lines of (a) equal intensity and (b) different intensity.

(a) Since according to the Rayleigh criterion v_0 should correspond to the minimum of the function $I_1(v)$, obviously $v_0 = \pi$. In the case that $I_{01} = I_{02}$ (Fig. 5.28a):

$$I_{t}(0) = I_{01} = I_{t}(v_{0}) = I_{t \max};$$
$$I_{t}(v_{0}/2) = I_{t}(\pi/2) = I_{01}\left(\frac{4}{\pi^{2}} + \frac{4}{\pi^{2}}\right) = 0.81I_{01} = I_{t \min}$$

and therefore the contrast C = 0.19.

(b) Let the intensity in the center of the first line be greater than that of the second line and their positions be the same as in (a) above (the minimum of the first line coincides with the maximum of the second one). Denoting the ratio $I_{01}/I_{02} = q$, we obtain from the previous expression for total intensity

$$I_{t}(0) = I_{01};$$
 $I_{t}(v_{0}) = I_{02};$ $I_{t}(v_{0}/2) = I_{t}(\pi/2) = I_{02}\left(\frac{4q}{\pi^{2}} + \frac{4}{\pi^{2}}\right) = I_{t\min}$

Calculation of the limiting contrast *C* which is equal to 0.05 yields $1 - C = 0.95 = I_{t \min}/I_{02} = 4(1+q)/\pi^2$, and therefore q = 1.344. This result means that if the intensity of the second line is less than 74.4% of that of the first one, the two lines cannot be resolved (registered as separated) although they are positioned according to the Rayleigh criterion.

5.12. For a prism with refraction angle β and refractive index *n* the minimum deviation angle φ_{\min} was derived in Problem 1.17 where we found

$$\varphi_{\min} = 2 \arcsin\left(n \sin\frac{\beta}{2}\right) - \beta.$$

From this expression one can get $\sin[(\varphi_{\min} + \beta)/2] = n \sin(\beta/2)$. Differentiating with respect to λ gives

$$\frac{1}{2}\cos\left(\frac{\varphi_{\min}+\beta}{2}\right)\frac{\mathrm{d}\varphi_{\min}}{\mathrm{d}\lambda} = \frac{\mathrm{d}n}{\mathrm{d}\lambda}.\sin\frac{\beta}{2}$$

from which we finally get

$$\frac{\mathrm{d}\varphi_{\min}}{\mathrm{d}\lambda} = \frac{\mathrm{d}n}{\mathrm{d}\lambda} \times \frac{2\sin(\beta/2)}{\sqrt{1 - n^2\sin^2(\beta/2)}}.$$

This is the expression for the angular dispersion of the prism working around the direction of minimal deviation, φ_{\min} .

5.13. (a) Since the refractive index difference, $\Delta n = n_F - n_C$, is known for two wavelengths, denoted as F and C, we use these wavelengths, $\lambda_C = 656.27$ nm; and $\lambda_F = 486.13$ nm, in order to estimate the dispersion power of the prism made of BK-7 glass:

$$\frac{\mathrm{d}n}{\mathrm{d}\lambda} = \frac{\Delta n}{\Delta\lambda} = \frac{0.008054}{170.14 \,\mathrm{nm}} = 4.734 \times 10^{-5} \,\mathrm{nm}^{-1}.$$

Then we use Eq. (5.16) to find the resolution (dimensionless):

$$\Re = \frac{\mathrm{d}n}{\mathrm{d}\lambda}B = 4.734 \times 10^{-5} \times 30 \times 10^{6} = 1,420$$

and therefore the minimum resolvable spectral interval at a representative wavelength of 500 nm in the visible is $\delta \lambda = 500/1, 420 = 0.352$ nm.

(b) As the optics is supposed to be diffraction limited, we should find the maximum size of the aperture defining the diffraction limit of the system. From the geometry of rays refracted by the prism (see Fig. 5.29) we find

$$D_{\text{dif}} = AK \cos i = \frac{AM \cos i}{\sin \beta/2} = \frac{B \cos i}{2 \sin \beta/2}.$$

Since for minimum deviation the incident angle $i = \arcsin(n \sin 20^\circ) = 31.24^\circ$, we can calculate $D_{\text{dif}} = 30 \cos(31.24^\circ)/2 \sin(20^\circ) = 37.5$ mm. This value is less



FIGURE 5.29 Problem 5.13 – Refraction of rays in a prism.

than the diameter of the lens (40 mm), and hence the prism and not the lens will cause the diffraction limit. Then the optimal size of the entrance slit can be found from Eq. (5.15):

$$b = \frac{0.5}{37.5}200 = 2.67 \,\mu\mathrm{m}$$

(of course, this is a theoretical value for the system with no aberration).

(c) If the stop of 30 mm is set in front of the lens then the actual diameter of the beam passing through the optics is limited by the stop and not by the prism. Therefore

$$D_{\text{dif}} = 30 \text{ mm}; \quad B_{\text{eff}} = \frac{2 \times 30 \times \sin 20^{\circ}}{\cos 31.24^{\circ}} = 24 \text{ mm}; \quad \Re = 1,136;$$

$$\delta \lambda = 0.44 \text{ nm}; \quad b = \frac{0.5}{30} 200 = 3.33 \text{ }\mu\text{m}.$$

If the entrance slit remains as in (b) above the spectrum will be of lower resolution (lower contrast) and all spectral lines will be of significantly lower intensity.

5.14. (a) As in Problem 5.13 we find first the dispersion power of the prism made of BK-7 glass: $dn/d\lambda = 4.734 \times 10^{-5}$ /nm. We then use Eq. (5.16) to calculate the system resolution: $\Re = 60 \times 10^6 \times 4.734 \times 10^{-5} = 2,840$.

(b) We find the wavelength difference between the spectral line of the source and that of the pollutant. For the source we have $\lambda_1 = (10,000/16,800) \times 10^3 = 595.24$ nm and for the pollutant $\lambda_2 = (10,000/16,790) \times 10^3 = 595.59$ nm, so that $\Delta \lambda = 0.35$ nm. Now we calculate the angular dispersion of the prism, D_{φ} , using the analytical expression found in Problem 5.12:

$$D_{\varphi} = 4.734 \times 10^{-5}$$
. $\frac{2 \sin 30^{\circ}}{\sqrt{1 - 1.5163^2} (\sin 30^{\circ})^2} = 7.26 \times 10^{-5} / \text{nm}$

Hence, the linear dispersion in the output plane is $D_l = D_{\varphi} f' = 7.26 \times 10^{-5} \times 300 \times 10^3 = 21.78 \,\mu$ m/nm, which yields for two lines at 0.35 nm the separation $\Delta l = 21.78 \times 0.35 = 7.6 \,\mu$ m. This value is less than a single pixel (10 μ m) of the detector array. Therefore, in this case the pollutant cannot be detected.

(c) Replacing the dispersion element by the prism made of SF-5 glass will increase the dispersion power and the line separation in the output plane. We obtain in this case

$$\frac{\mathrm{d}n}{\mathrm{d}\lambda} = \frac{0.020884}{170.14} = 12.27 \times 10^{-5}/\mathrm{nm};$$

$$D_{\varphi} = \frac{\mathrm{d}n}{\mathrm{d}\lambda} \cdot \frac{2\sin 30^{\circ}}{\sqrt{1 - 1.6727^{2}(\sin 30^{\circ})^{2}}} = 22.38 \times 10^{-5}/\mathrm{nm};$$

$$D_{l} = 67.14 \,\mu\mathrm{m/nm}; \quad \Delta l = 67.14 \times 0.35 = 23.5 \,\mu\mathrm{m}.$$

This means that the two spectral lines are separated by more than twice the pixel size and the sampling requirements of the Nyquist theorem are satisfied, so that the pollutant can be detected in such a configuration.

5.15. Find the spectral interval corresponding to the minimum spot in the output plane:

$$\delta\lambda = \frac{\Delta l}{dl/d\lambda} = \frac{30}{21.78} = 1.377 \text{ nm}; \quad \Re = \frac{595}{1.377} = 432.$$

Here it is assumed that the prism is made of BK-7 glass (see Problem 5.14). Since $\delta\lambda > 0.35$ nm, the system cannot reveal the pollutant mentioned in Problem 5.14. Replacing the prism by another one made of SF-5 glass still cannot solve the problem: $\delta\lambda = 30/67$. 14 = 0.447 nm > 0.35 nm.

5.16. In considering the OPD between two rays 1 and 2, Δ_{21} , we will address it as a delay of the second ray with regard to the first and we will also assume that ray 1 is always located to the left of ray 2. We should also keep in mind that the incident angle ψ obeys the regular sign conventions described in Section 1.1 (it is positive if the optical axis, or the normal to the grating, should be rotated clockwise in order to coincide with the ray and it is negative if rotation is counterclockwise) whereas the angle of diffraction, φ , concerns the rays after reflection and therefore the signs are opposite (the angle is negative if the normal is rotated clockwise and it is positive if rotation is counterclockwise). Furthermore, we will divide the whole OPD into two parts, $\Delta_{21} = \Delta_{21}^{(i)} + \Delta_{21}^{(d)}$, where the first is related to the incident beam and the second describes the rays after reflection (diffraction).

(a) The beam incident on the grating from the right, as depicted in Fig. 5.18a. In this case $\psi > 0$ and $\Delta_{21}^{(i)} = BC = -d \sin \psi$ is negative since ray 2 reaches the grating earlier than ray 1 ("negative" delay). For diffraction directions left of the normal $\varphi > 0$ and the delay $\Delta_{21}^{(d)} = AD = d \sin \varphi$ is positive (ray 2' passes further than ray 1'). Hence, the overall delay is $\Delta_{21} = d(-\sin \psi + \sin \varphi)$. For diffraction directions right of the normal $\varphi < 0$ and the corresponding delay $\Delta_{21}^{(d)}$ is also negative, so that the same expression for Δ_{21} as above describes correctly the total OPD.

(b) The beam incident on the grating from the left. In this case $\psi < 0$, but the delay $\Delta_{21}^{(i)} = BC = -d \sin \psi$ is positive (ray 2 is behind ray 1). For diffraction directions right of the normal $\varphi < 0$ and the corresponding delay $\Delta_{21}^{(d)}$ is also negative. The overall OPD is positive for diffraction angles smaller than ψ and is negative if $|\varphi| > |\psi|$. For diffraction directions left of the normal $\varphi > 0$ and the delay $\Delta_{21}^{(d)} = AD = d \sin \varphi$ as well as the overall OPD are positive and again the same expression for Δ_{21} remains valid.

5.17. From Eq. (5.21) one obtains

$$\sin\varphi_{\lambda\,\max}^{(m)} = \sin\psi + \frac{m}{d}\lambda_{\max}; \quad \sin\varphi_{\lambda\,\min}^{(m+1)} = \sin\psi + \frac{m+1}{d}\lambda_{\min}.$$

Subtracting the second expression from the first we obtain the (angular) difference between two corresponding directions. The overlapping starts when this difference is equal to zero, which yields

$$\frac{m(\lambda_{\max} - \lambda_{\min})}{d} = \frac{\lambda_{\min}}{d} \quad \text{or} \quad \Delta \lambda = \frac{\lambda_{\min}}{m}$$

The greater the working diffraction order the smaller the spectral interval available without overlapping.

5.18. We will perform the calculation for the wavelength $\lambda = 5,000$ Å and start from Eq. (5.29) where we first assume $\cos \varphi = 1$ in order to estimate the total number of grooves, N, of the grating. Taking into account that the required resolution of the system is $\Re = 5,000$ Å/0.2 Å = 25,000, we get $N = \Re/m = 12,500$ which enables one to find the period, d, of the grating: d = 25 mm/12,500 = 2 μ m. Therefore, the grating spatial frequency is 500 lp/mm. The parameters of the grating grooves can be obtained from Eqs. (5.24) and (5.25). We rewrite the first one using the data of the problem as follows: $2\cos(-15^\circ - \gamma) \times \sin(-\gamma) = p$, where $p = m\lambda/d = -0.5$; and furthermore: $\cos 15^\circ \sin(2\gamma) - 2\sin 15^\circ \sin^2(\gamma) = -p$. Replacing $\sin(2\gamma)$ and $2\sin^2(\gamma)$ by appropriate expressions with $\tan \gamma$ in a standard way and denoting $x = \tan \gamma$ we get the following second order equation with regard to x: $x^2(p - 2\sin 15^\circ) + 2x\cos 15^\circ + p = 0$. This gives $x = \tan \gamma = 0.3092$; $\gamma = 17.18^\circ$. With this value of the groove inclination angle we get from Eq. (5.25) the active size of the groove mirror:

$$b = \frac{0.5}{2\cos 15^\circ \times \sin 17.18^\circ} = 0.876 \,\mu\text{m}.$$

Finally, using Eq. (5.21) we calculate the diffraction angle where most energy is concentrated: $\sin \varphi_{\text{max}}^{(-2)} = -\sin 15^{\circ} - 0.5$; $\varphi_{\text{max}}^{(-2)} = 49.36^{\circ}$. This also yields the corrected value of the system resolution: $\Re = (2 \times 12, 500)/\cos 49.36^{\circ} = 38,385$.

5.19. The grating period is $d = 1/300 \times 10^3 = 3.33 \,\mu\text{m}$. Substituting this value in Eq. (5.21) which describes the conditions of diffraction for the principal maxima and keeping in mind that in autocollimated architecture the incident beam is parallel to the optical axis and therefore the tilted angle is the incident angle ψ of the grating, we draw the conclusion that the system is capable of operating in different diffraction orders. As is evident from Eq. (5.29), the higher the order *m* the greater the resolution. However, this is true only if there is no vignetting in the diffracted



FIGURE 5.30 Problem 5.19 – Geometry of diffracted rays between a grating and lens.

light (all rays leaving the grating and participating in the production of the spectrum pass through the lens L). The geometrical consideration of this requirement is demonstrated in Fig. 5.30. As we see, in order to avoid vignetting the following limiting condition should be satisfied: $y_1 + y_2 < D/2$; or in terms of the system parameters:

$$s\tan(\varphi+\psi) + \frac{G}{2}\frac{\cos\varphi}{\cos(\psi+\varphi)} < \frac{D}{2}$$

where *G* and *D* are the grating size and the lens diameter, respectively. Besides this, it is understandable from the system architecture that the detector array should be positioned above the optical axis and therefore the useful diffraction beams propagate only upward, meaning that $(-\varphi) < \psi$.

(a) Calculation of φ for different *m* can be done with Eq. (5.21). The results show that only two values, m = -1 and m = -2, obey both limiting conditions: for the first we get $\varphi = 4.52^{\circ}$; $(y_1 + y_2) = 48.65$ mm; and for the second $\varphi = -5.81^{\circ}$; $(y_1 + y_2) = 28.77$ mm. For m = -3, m = -4, etc., the diffraction beams are directed downward and cannot be accepted. Therefore, the optimal order is m = -2. The system resolution found from Eq. (5.29) becomes $\Re = 2 \times 300 \times 25.4 = 15,240$ and the minimum resolvable spectral interval is $\delta \lambda = 600/15,240 = 0.04$ nm. Exploiting the results of Problem 5.17, we find the free spectral interval without overlapping: $\Delta \lambda = 600/2 = 300$ nm.

(b) The location of the detector array is related to the angle $\varphi_{\text{max}}^{(-2)}$ of the chosen principal maximum. In our case it is -5.81° , so that the first pixel of the array should be positioned at a height $H = f' \tan(\psi + \varphi_{\text{max}}^{(-2)}) = 1,200 \times \tan(15^{\circ} - 5.81^{\circ}) = 194.1$ mm above the optical axis. The angular and linear dispersion of the system are calculated from Eqs. (5.26) and (5.27):

$$D_{\varphi} = \frac{2}{3.33 \times 10^3 \times \cos 5.81^{\circ}} = 6 \times 10^{-4} / \text{nm}; \quad D_l = D_{\varphi} f' = 0.72 \text{ mm/nm}.$$
Hence, the minimum resolvable spectral interval needs in the output plane the segment $\Delta l = D_l \times \delta \lambda = 0.72 \times 10^3 \times 0.04 = 28.8 \,\mu\text{m}$. Therefore, the size of a single pixel should be half of this value, i.e., 14.4 μ m, and the total number of pixels in the detector array should be $M = (300/0.04) \times 2 = 15,000$ (if registration of all possible wavelengths in the full spectral range is desirable; usually the detector array is much shorter – very seldom does M exceed 4,000).

5.20. We find the period d of the grating $d = 1000/1, 200 = 8.333 \,\mu\text{m}$ and proceed to Eq. (5.21) for the diffraction angle of the first principal maximum: $\sin \varphi_{\text{max}}^{(1)} = \sin 10^{\circ} + 0.5/0.8333 = 0.7736$; $\varphi_{\text{max}}^{(1)} = 50.68^{\circ}$. Then the linear dispersion is

$$D_l = D_{\varphi} f' = \frac{f'}{d \cos \varphi_{\text{max}}^{(1)}} = \frac{150}{0.833 \cos 50.68^\circ} = 284 \,\mu\text{m/nm}$$

and finally $\Delta \lambda = b'/D_l = 20/284 = 0.0704$ nm.

5.21. Theoretically the resolution of the grating in the problem is high enough $(\Re = mN = 15, 240 \text{ even for the first diffraction order which potentially enables one to register spectral variations as small as 830 nm/15, 240 = 0.056 nm). However, since the laser illuminates only a portion of the grating this increases the minimum diffraction spot achievable in the output plane of the spectrometer. The corresponding diffraction angle of this spot is$

$$\Delta \theta = \frac{\lambda}{Nd} = \frac{0.83}{2 \times 10^3} = 4.15 \times 10^{-4}$$

and all the wavelengths propagating inside of this angle cannot be separated from one another. This defines the real spectral resolution, $\delta\lambda$, achievable in the system. As the angular dispersion of the grating $D_{\varphi} = m/d \cos \varphi \approx 6 \times 10^{-4}$ m, one can estimate the spectral uncertainty as follows: $\delta\lambda = \Delta\theta/D_{\varphi} = 0.69/m$ nm. Therefore, if the spectrometer is operated at the first diffraction order the minimum width of a spectral line is 0.69 nm; if the instrument is set for m = 2 the width is 0.345 nm, etc.

It is well known that laser diode instability might be significant. Due to the variation of temperature, for example, the wavelength of the laser diode might increase 1 nm for every 5°C. To be able to reveal such variations the testing spectrometer should have at least twice as high a spectral resolution. This means that the spectrometer mentioned in the problem does not suit the study if it is set for the first diffraction order – it has to be operated at m = 2 at least.

5.22. With the grating period $d = 1,000/2,400 = 0.417 \ \mu\text{m}$ we get from Eq. (5.21) $\sin \varphi_{\text{max}}^{(-1)} = \sin 30^{\circ} - (0.3/0.417) = -0.2194$ and therefore the direction to the exit slit is determined by the angle $\varphi_{\text{max}}^{(-1)} = -12.68^{\circ}$. As this angle

is negative, it means that both the entrance slit and the exit slit are positioned on the Rowland circle on the same side with regard to its diameter normal to the grating. The distances to the slits are calculated from Eq. (5.31) which yields $r = 500 \cos 30^\circ = 433$ mm; and $r' = 500 \cos 12.68^\circ = 488$ mm. As a result, the optical magnification of the grating, V (see Eq. (5.32)), is equal to 1.1266 and the exit slit is 0.225 mm.

5.23. (a) The entrance slit as well as the center of the CCD line detector should be positioned on the Rowland circle. From the first relation of Eq. (5.31) we find the illumination angle ψ : $\cos \psi = 0.5$; $\psi = 60^{\circ}$. Now we can proceed to Eq. (5.21) to find the angle of diffraction:

$$\sin\varphi_{\max}^{(-2)} = \sin 60^{\circ} - \frac{2 \times 0.3}{1,000/600} = 0.506; \ \varphi_{\max}^{(-2)} = 30.4^{\circ}.$$

The second relation of Eq. (5.31) gives $r' = 250 \cos 30.4^{\circ} = 215.6$ mm. This is the distance to the CCD detector center along the segment r'. The detector itself should be perpendicular to the segment r'.

(b) Using Eq. (5.33) we find the linear dispersion of the grating:

$$D_l = \frac{m\rho}{d} = \frac{2 \times 250}{1,667} = 300 \,\mu$$
m/nm.

Hence, the minimum resolvable spectral interval is $\delta \lambda = 2s/D_l = (2 \times 15)/300 = 0.10$ nm and the total spectral range is $\Delta \lambda = 1,024 \times 0.1 = 102.4$ nm.

5.24. If not specified, the incident radiation is supposed to be normal (perpendicular) to the filter surfaces. Therefore, using r = 0 in Eq. (5.35) one can find the wavelength for which the interference maximum will occur: $\lambda = \Delta = 2tn = 2 \times 0.2 \times 1.4 = 0.56 \,\mu\text{m}$. Then we obtain from Eq. (5.39)

$$m = 1;$$
 $N_{\rm e} = \frac{\pi\sqrt{0.95}}{1 - 0.95} = 61$

and finally from Eq. (5.38) FWHM = $\delta \lambda = 0.56/61 = 9$ nm.

5.25. Tilting of the filter with regard to the incident radiation will change the optimal wavelength and, slightly, the FWHM. Indeed, proceeding as in Problem 5.24, but now taking into account that $r \neq 0$, we obtain for the incident angle 20° sin $r = \sin 20^{\circ}/1.4 = 0.2443$; $r = 14.1^{\circ}$. Expression (5.35) gives $\lambda = 2 \times 0.2 \times 1.4 \times \cos 14.1^{\circ} = 543$ nm. Since N_e remains the same as in Problem 5.24 we get for FWHM: $\delta \lambda = 543/61 = 8.9$ nm. A similar calculation for the tilting angle 30° yields $\lambda = 523$ nm; $\delta \lambda = 8.6$ nm.

5.26 We first estimate the effective number of rays, N_e , required for achieving a FWHM as small as 1 nm in a single cavity filter: $N_e = \lambda/\delta\lambda = 600 \ (m = 1 \text{ in an})$



FIGURE 5.31 Problem 5.27 – Spectral intensity distribution in a convergent beam.

ordinary interference filter). At high reflection, *R*, one may use the approximate expression $N_e \approx \pi/(1 - R)$. Hence, the necessary reflection should be $R = 1 - (\pi/600) = 0.995$ which is technologically impossible and this is the reason why a multi-cavity architecture is used if a very narrow bandpass (characterized by FWHM) is required.

5.27. The convergent beam mentioned in the problem can be considered as a combination of parallel beams of different incident angles, from 0° to $\pm 10^{\circ}$, each one contributing to the overall intensity distribution a fraction similar to that shown in Fig. 5.23b, but centered at the wavelength λ' dependent on the angle of incidence. Proceeding as in Problem 5.25 we get for the maximum angle of 10° : $\lambda' = 2tn \cos r = 2tn \cos[\arcsin(\sin 10^{\circ}/1.4)] = 496$ nm. Assuming that all tilted beams are of the same intensity as the central one, we find the total intensity distribution shown in Fig. 5.31, which gives FWHM = 9 nm.

5.28. The wavelengths which correspond to the spectral lines of the problem are $\lambda_1 = 3,100$ Å and $\lambda_2 = 3100.29$ Å and the resolution required is $\Re = 3,100/0.29 = 10,690$. For an air-spaced Fabry–Perot etalon we get from Eq. (5.35) $m\lambda = 2t$ which together with Eq. (5.41) gives

$$t = \frac{m\Re}{2}\delta\lambda = \frac{\Re^2}{2N_{\rm e}}\delta\lambda.$$

The effective number of rays depends on the reflectivity of the coating, according to Eq. (5.39), and therefore the air spacing is also related to *R*. For R = 0.85 we obtain from the above expression t = 0.8 mm. For R = 0.90, 0.92, and 0.95 we obtain t = 0.534 mm, 0.427 mm, and 0.269 mm, respectively.

5.29. (a) We rewrite Eq. (5.35) for the first (on-axis) maximum to calculate the interference order:

$$2tn = m\lambda; \quad m = \frac{2 \times 1.7 \times 1.5}{0.5 \times 10^{-3}} = 10,200.$$

For reflectivity R = 0.95 Eq. (5.39) yields $N_e = 61$. By substituting this value in Eq. (5.38) we find $\delta \lambda = 8 \times 10^{-3}$ Å and therefore $\Re = (5,000/8 \times 10^{-3}) = 625,000$.

(b) We use Eq. (5.40) to find the radii of the first and the second rings:

$$\rho_1 = 200\sqrt{\frac{2}{10,200}} = 2.8 \text{ mm}; \quad \rho_2 = 200\sqrt{\frac{4}{10,200}} = 3.96 \text{ mm}$$

(c) As the output aperture S' is of 3.5 mm radius, the second ring of the wavelength 0.5 μ m is already outside of S'. The wavelength $\overline{\lambda}$ for which the second ring is still observed can be calculated from Eq. (5.40) as follows:

$$3.5 = 200 \times \sqrt{\frac{2\bar{\lambda}n}{1.7}}; \quad \bar{\lambda} = 0.174 \,\mu\mathrm{m}$$

which is out of the visible region. Therefore, any wavelength from the visible region can be represented in S' by a single ring only.

5.30. We perform calculations for both limits of the working interval. For the smallest wavelength, using the definition of wavenumber one can obtain

$$\delta N = \frac{\delta \lambda}{\lambda^2} = \frac{2 \times 10^{-8}}{10^{-8}} = 2 \text{ cm}^{-1}.$$

Then, by substituting this value in Eq. (5.48) we get the range of the scanning mirror movement as $l = 1/(2 \times \delta N) = 0.25$ cm. A similar procedure for the largest wavelength gives

$$\delta N = \frac{2 \times 10^{-8}}{25 \times 10^{-8}} = 0.08 \text{ cm}^{-1}; \quad l = \frac{1}{2 \times 0.08} = 6.25 \text{ cm}$$

5.31. From the data of Problem 5.30 we draw the conclusion that a resolution of 0.2 nm through the whole working spectral interval of 4 μ m requires 2 × 4,000/0.2 = 40,000 sampling points (we also take into account the sampling conditions according to the Nyquist theorem). Since the detector bandwidth is 20 kHz, the sampling in the time domain can be performed at time intervals $\tau = 1/(2 \times 20 \times 10^3) = 25 \,\mu$ s. For the FFT digital algorithm the number of sampling points in the time domain and in the frequency domain should be the same, so the full time of scanning is $T = 25 \times 10^{-6} \times 40,000 = 1$ s which requires for a scanning range of 6.25 cm (found in Problem 5.30) a scanning velocity of V = 6.25 cm/s.

5.32. First we should find the absorption factor of the silver particles. From Eq. (5.11) we have

$$\alpha_{\rm Ag} = \frac{4\pi \times 3.64}{0.59 \times 10^{-3}} = 7.75 \times 10^4 \, {\rm mm}^{-1}.$$

Then, by substituting this value in Eq. (5.51) as well as the ratio of the detector signals, $T_{21} = 0.05$, one can obtain the absorption factor of the media:

$$\alpha_{\rm M} = \frac{-\ln(0.05)}{(3.5 - 1.0)} = 1.198 \, {\rm mm}^{-1}$$

and finally the concentration C of the silver particles in the solution (see Eq. (5.49)):

$$C = \frac{1.198}{7.75 \times 10^4} = 1.546 \times 10^{-5} = 15.46$$
 ppm (parts per million).

Non-contact Measurement of Temperature

6.1. Thermal Radiation Laws and Surface Properties

Thermal radiation is a collection of electromagnetic waves emitted by a substance and results from the random motion of microparticles (atoms, molecules, ions, and electrons) constituting the radiating body. The motion of microparticles is a characteristic of matter and it occurs at any time and anywhere in all substances, either in the condensed phase (solids or liquids) or in the gaseous forms. If a moving particle has an electric charge that moves (randomly oscillates or randomly jumps from point to point) together with the particle mass, such a motion is inevitably accompanied by the generation of electric and magnetic fields propagated as waves in all directions. Since the motion of one particle is not correlated with that of others, the emitted electromagnetic waves are also not correlated with each other, but radiate spontaneously, in a chaotic manner. Thus, thermal radiation is completely different from the "well-organized," coherent light of lasers and different from the radiation of other sources converting electrical, biological, or chemical energy into emitted electromagnetic waves (some of them are described in Chapter 3).

As the thermodynamic temperature, T, is the basic measure of the kinetic energy of randomly moving particles, it is quite understandable that thermal radiation is governed primarily by the temperature of a radiating substance. Measuring thermal radiation allows one to estimate the temperature of a thermal source, once the relation between the temperature and radiation power is known and well established. However, as will be explained later, radiation depends not only on Tbut also on other properties of a radiating body.



FIGURE 6.1 External radiation incident on a surface.

To explain the laws of thermal radiation we first discuss a very simple and general case when radiation of an external source is incident on a surface separating two media: the first being fully transparent (it could also be a vacuum) from which the radiation comes and the second being the body under consideration (see Fig. 6.1). As is well known, some of the incident energy is reflected back into the first medium while the rest is propagated inside the body and finally fully absorbed there (if the body is semi-infinite) or emerges on the other side of it. Considering the spectral values related to a specific wavelength λ and denoting the incident energy as E_{λ} and the reflected and absorbed energies E'_{λ} and E''_{λ} , respectively, one obtains from the energy conservation law

$$E_{\lambda} = E_{\lambda}' + E_{\lambda}''; \quad 1 = \frac{E_{\lambda}'}{E_{\lambda}} + \frac{E_{\lambda}''}{E_{\lambda}} = R_{\lambda} + A_{\lambda}$$
(6.1)

where the spectral reflectance R_{λ} and the spectral absorptance A_{λ} are introduced. Furthermore, assuming that the second medium (the body) has temperature *T* and emits thermal radiation characterized by spectral emittance $e_{\lambda}(T)$, we can express Kirchhoff's law as follows:

$$\frac{e_{\lambda}(T)}{A_{\lambda}(T)} = \text{const} = e_{B\lambda}(T)$$
(6.2)

which states that the ratio of spectral emittance to spectral absorptance is a universal function for all bodies (all thermal sources). This ratio varies when the temperature and/or chosen wavelength vary, but it does not depend on the body material. The universal function $e_{B\lambda}(T)$ in Eq. (6.2) is obviously the emittance of a substance for which $A_{\lambda} = 1$. Such a body is called a black body and it has the maximum possible emittance for a given wavelength and temperature. The radiation properties of real surfaces are usually defined relative to this maximum emittance by the value ε_{λ} called emissivity:

$$\varepsilon_{\lambda}(T) = \frac{e_{\lambda}(T)}{e_{B\lambda}(T)}.$$
(6.3)

By substituting Eq. (6.3) in Eq. (6.2) and then in Eq. (6.1) yields

$$\varepsilon_{\lambda} = A_{\lambda} = 1 - R_{\lambda} \tag{6.4}$$

and this expression is strictly valid for any surface (if all three values are related to the same temperature T and the same wavelength λ , of course). It should be noted, however, that sometimes the emissivity, reflectance, and absorptance used in practice are determined as averaged values over a rather wide spectral interval, or even over all wavelengths in some cases. Then the last equality is no longer valid and the averaged values cannot be related to each other as simply as in Eq. (6.4), but should be estimated separately.

The universal function $e_{B\lambda}(T)$ appearing in Kirchhoff's law is defined by Planck's law:

$$e_{\rm B}(\lambda, T) = \frac{C_1 \lambda^{-5}}{\exp\left(-\frac{C_2}{\lambda T}\right) - 1}$$
(6.5)

where $C_1 = 3.740 \times 10^{-16}$ W m² and $C_2 = 1.4387 \times 10^{-2}$ m K are universal constants and Eq. (6.5) is related to the radiation power emitted from a surface of 1 m² in all directions of a hemisphere (2π sr). This function is shown in Fig. 6.2 for three different temperatures. As can be seen, the wavelength of maximum emission (λ_{max}) moves to shorter wavelengths as temperature increases. The relation between *T* and λ_{max} is governed by Wien's law (it can also be derived directly from Eq. (6.5)):

$$\lambda_{\max} T = 2,898 \,\mu \mathrm{m} \,\mathrm{K}.$$
 (6.6)



FIGURE 6.2 Black body radiation function (Planck's function) at different temperatures $(T_1 > T_2 > T_3)$.

If we are interested in radiation emitted by a body at all wavelengths, Eq. (6.5) should be integrated from zero to infinity, which results in the following formula:

$$e_{\rm B}(T) = \int_{0}^{\infty} e_{\rm B}(\lambda, T) d\lambda = \sigma T^4$$
(6.7)

where $\sigma = 5.668 \times 10^{-8} \text{ Wm}^{-2} \text{ K}^{-4}$ is the Boltzmann constant and Eq. (6.7) is known as the Stefan–Boltzmann law.

The radiation properties of radiating surfaces are characterized by emissivity, ε . Since these properties depend not only on wavelength and temperature but also on the surface shape and roughness, it is common practice to introduce the angular spectral emissivity defined as the following:

$$e_{\lambda,T}(\theta,\varphi) = \frac{I_{\lambda}(\theta,\varphi)}{I_{\rm B}(\lambda,T)}$$
(6.8)

where instead of hemispherical characteristics e_{λ} and e_{B} the radiation intensities are used. Intensity is defined as the radiation power radiated by a unit surface perpendicular to the direction of observation (defined by angular coordinates θ, φ) in a solid angle of 1 sr around the direction of observation. Sometimes such values are termed surface brightness. A black body surface has constant brightness independent of observation direction (Lambert's law) and this is the reason why angular coordinates in Eq. (6.8) appear in the numerator only. If a black body surface is of constant area *S* and located in a plane which is not changed (not turned together with the observation angle), then intensity I_{B} measured at different angles θ varies as $\cos(\theta)$ (according to the projection of *S* on the plane perpendicular to the observation).

Examples of directional emissivity of two surfaces, dielectric and metal, are shown in Fig. 6.3. As can be seen, at small observation angles (up to about 45°) both surfaces behave as lambertian surfaces (emissivity is almost constant). However, at greater angles the emissivity of the dielectric decreases monotonically while for the metal it increases and only at very large angles does it decrease rapidly. Data on the emissivity of different surfaces are presented in Appendix 4.

Problems

6.1. An optical system comprising a collection lens (diameter 30 mm, focal length 50 mm) and a silicon detector of 3 mm in size is set in a furnace window at a distance of 5 m from the furnace inner wall for measuring the wall temperature. The detector responsivity is 0.28 A/W and its maximum current is 1 mA.



FIGURE 6.3 Spatial distribution of emissivity of a radiating surface: (a) a dielectric material; (b) a metal.

- (a) Assuming that the emissivity of the wall $\varepsilon = 0.8$ and its temperature $T = 1,750^{\circ}$ C, find the attenuation of a neutral density filter which should be inserted in front of the system if the spectral interval of the detector sensitivity is from 0.4 to 1.1 μ m and normal measurement is to be done in the middle of the system dynamic range. [Note: The emissivity of the wall can be assumed to be spectrally independent.]
- (b) How will the detector reading be changed when the system is moved aside and its optical axis is tilted 60° to the wall normal?

6.2. In the course of temperature measurement of a glass sheet in a tin bath the measurement assembly, located 1 m above the glass, is slightly rotated in order to measure the temperature at several points (A, B, and C) along the sheet. The distances AB = BC = 70 cm. Assuming that the refractive index of the glass is n = 1.5 and the glass temperature is about 500°C, find the measurement uncertainty due to rotation of the system.

6.3. It is well known that classic thermodynamics failed in predicting the spectral behavior of black body radiation at short wavelengths ($\lambda \rightarrow 0$) and instead of Planck's law (Eq. (6.5)) the classic approach resulted in Wien's formula:

$$e_{\rm B}(\lambda, T) = C_1 \lambda^{-5} \exp\left(-\frac{C_2}{\lambda T}\right)$$
 (6.9)

where the constants are the same as in Eq. (6.5). This expression is still very useful, due to its simplicity, in many applications. Keeping in mind that in most practical cases temperature varies from 300 K to 3,000 K and assuming that at each temperature the useful spectral range is $\lambda_{max}/2 < \lambda < 2\lambda_{max}$, estimate the accuracy of Wien's formula relative to the precise expression (Eq. (6.5)).



FIGURE 6.4 Problem 6.4 – Model of a black body.

6.4. A good model of a black body is a hollow sphere with a small aperture (see Fig. 6.4). Due to the geometry of the model the aperture absorbs practically any incident beam so that absorptance is very high. On the other hand, emittance remains practically independent of the reflectivity of the wall material because of multiple reflections inside the sphere. Assuming the size of the aperture is 5% of the sphere diameter and the wall is fully diffusive with reflectance R = 80%, estimate the accuracy of the model.

6.2. Optical Methods of Temperature Measurement

Optical instruments for temperature measurement (often called pyrometers) are based on universal laws of thermal radiation described in Section 6.1. As explained earlier, additional information required for the correct interpretation of measurement data is the emissivity of the surface under study. The basic configuration of a pyrometer is presented in Fig. 6.5. The device comprises radiation collecting optics, L, and a detector, D. In some methods (see below) an optical filter F (shown by dotted lines) is inserted in front of the detector. In general, the measuring process consists of two steps: measurement of radiation of the studied body, M, and a calibration procedure with a black body model, usually equipped with electrical



FIGURE 6.5 Basic configuration of an optical pyrometer: (a) measurement of radiation emitted by an object being studied; (b) calibration with a black body model.

measures enabling one to control and to change its temperature, $T_{\rm B}$. Of course, calibration can be done in advance, from time to time, independent of the measurement of the body being studied. What is important is that all the geometrical parameters, distances and collecting angle ω , be equal in the measurement and in the course of calibration.

Three kinds of temperature are usually defined and correspondingly three different approaches are commonly used. We will consider all of them, keeping in mind that in any case our goal is to find the true (thermodynamic) temperature, T.

Radiation Temperature

Let a detector be capable of registering radiation in a wide spectral range (theoretically at all wavelengths) and let the integral emissivity of a body (averaged over all wavelengths) be ε . Then the radiation power incident on the detector during the measurement and originating in emission of the segment dS of the tested object M is defined by the Stefan–Boltzmann law (Eq. (6.7)) as $E_1 = \varepsilon \sigma t^4 \omega \, dS$. While calibrating with a black body, the power registered at each temperature $T_{\rm B}$ is equal to $E_2 = \sigma T_{\rm B}^4 \omega \, dS$. The radiation temperature of the body M is defined as the temperature $T_{\rm BR}$ of the black body that gives an optical power equal to that registered with the body M, i.e., $E_1 = E_2$, and therefore

$$\sigma T_{\rm BR}^4 = \varepsilon \sigma T^4; \quad T = \frac{T_{\rm BR}}{\sqrt[4]{\varepsilon}}.$$
 (6.10)

Obviously our aim is the true temperature, T. As we see, once the radiation temperature, T_{BR} , is found and emissivity is known the true temperature, T, of body M is calculated from Eq. (6.10).

Color Temperature

A device registers radiation in two wavelengths, λ_1 and λ_2 , and the corresponding spectral emissivities, ε_1 and ε_2 , of a studied surface are known. Also, the process is performed separately for each wavelength by using appropriate narrow-band filters F_1 and F_2 in both measurement and calibration processes. Consider the ratio Qof registered data at these two wavelengths and define the color temperature of the studied body M as the temperature of a black body, T_{BC} , which yields a value Q equal to that obtained in measurements with the object M. According to this definition one obtains

$$\frac{\varepsilon_1 e_{\rm B}(\lambda_1, T)}{\varepsilon_2 e_{\rm B}(\lambda_2, T)} = \frac{e_{\rm B}(\lambda_1, T_{\rm BC})}{e_{\rm B}(\lambda_2, T_{\rm BC})}.$$
(6.11)

By substituting in Eq. (6.11) Planck's formula (Eq. (6.5)) we get the non-linear algebraic equation with regard to *T*. This strict result may be further simplified if

we take into account the consideration made in Problem 6.3 and replace Planck's expression by Wien's formula (Eq. (6.9)) which is a good approximation for many cases, as shown earlier. Then we have

$$\frac{\varepsilon_1 C_1 \lambda_2^{-5} \exp\left(-\frac{C_2}{\lambda_1 T}\right)}{\varepsilon_2 C_1 \lambda_1^{-5} \exp\left(-\frac{C_2}{\lambda_2 T}\right)} = \frac{C_1 \lambda_2^{-5} \exp\left(-\frac{C_2}{\lambda_1 T_{BC}}\right)}{C_1 \lambda_1^{-5} \exp\left(-\frac{C_2}{\lambda_2 T_{BC}}\right)}$$

Finally, after taking the logarithm of both sides of the equation:

$$\frac{1}{T} = \frac{1}{T_{\rm BC}} - \frac{\ln(\varepsilon_1/\varepsilon_2)}{C_2 \left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right)}$$
(6.12)

In the last expression wavelengths are in micrometers, temperatures in Kelvin, and $C_2 = 14,388$.

As we see, the color temperature can be smaller or larger than the true temperature of a body M, depending on the ratio of the emissivities $\varepsilon_1/\varepsilon_2$ related to the two wavelengths of measurement.

Brightness Temperature

This case is similar to radiation temperature, but equalization of measured radiation quantities is performed for the spectral values related to a single wavelength, λ . Assuming again that the spectral emissivity ε_{λ} is known and defining the brightness temperature as the temperature T_{BS} of a black body at which the measured spectral radiation at calibration is equal to that of the measurement with a body M, we have

$$\varepsilon_{\lambda} e_{\rm B}(\lambda, T) = e_{\rm B}(\lambda, T_{\rm BS})$$
 (6.13)

which is again the non-linear algebraic equation with regard to T. Further approximation with Wien's formula (Eq. (6.9)) yields the following:

$$\varepsilon_{\lambda}C_{1}\lambda^{-5}\exp\left(-\frac{C_{2}}{\lambda T}\right) = C_{1}\lambda^{-5}\exp\left(-\frac{C_{2}}{\lambda T_{BS}}\right).$$

The final result is

$$\frac{1}{T} = \frac{1}{T_{\rm BS}} + \frac{\ln \varepsilon_{\lambda}}{C_2/\lambda}.$$
(6.14)

The last expression shows that the thermodynamic temperature T of a surface is always higher than its brightness temperature T_{BS} (since $\varepsilon_{\lambda} < 1$ in all cases).

Problems

6.5. The color temperature of a surface is measured with two wavelengths, $\lambda_1 = 0.45 \,\mu\text{m}$ and $\lambda_2 = 0.55 \,\mu\text{m}$. The spectral emissivity of the surface is 0.3 and 0.4, respectively.

- (a) Find the true temperature of the surface if the measured color temperature is 2,200 K.
- (b) Calculate the brightness temperature of the surface at both given wavelengths.

6.6. In a pyrometer of equal brightness intended for operation with the naked eye a filter F_1 (made of glass 20 mm thick with n = 1.6 and absorption factor $K = 0.5 \text{ cm}^{-1}$) is set in front of a light source S (see Fig. 6.6) while a narrow-band filter F_2 transparent for a wavelength of 0.5 μ m is positioned in front of the eye. If the emissivity of body M is $\varepsilon_{\lambda} = 0.5$ and the brightness temperature of source S is 2,850 K, what is the true temperature of the body?



FIGURE 6.6 Problem 6.6 – Configuration of a pyrometer of equal brightness.

6.7. The radiation temperature of the refractory wall in a furnace is 2,700°C. Calculate the heat flux emitted by the wall in the visible spectral interval, assuming the wall emissivity is 0.8 for all relevant wavelengths.

6.8. In manufacturing window glass by the float process, liquid glass is run out from a furnace onto the surface of liquid tin in a tin bath. The temperature of the tin surface is controlled at several points along the bath. Measurements are performed by an optical pyrometer where radiation is registered at two wavelengths, $\lambda_1 = 0.52 \,\mu\text{m}$ and $\lambda_2 = 0.45 \,\mu\text{m}$, and the resulting color temperature is found as 1,000°C near the furnace and 650°C near the bath outlet. Assuming the tin reflectance varies with wavelength as $R = 1 - (\text{const}/\sqrt{\lambda})$, find the difference between the measured values and the true temperature of the tin surface at both locations.

6.3. Measurement of Temperature Gradients

Optical methods for the measurement of gradients inside a test object are evidently concerned with materials transparent to radiation. We will consider here the method based on an interferometic configuration.

Measurement of gradients is a much more complicated task than the measurement of temperature in some predetermined locations. Generally speaking, gradients are revealed as a result of interferogram interpretation which takes into consideration that the shape of the interference fringes depends on the temperature and variation of the refractive index along the optical path. Since optical path differences (OPDs) which give the final interference pattern accumulate numerous local variations of n, it might occur that different spatial distributions of n(x, y, z)yield the same final result. From a mathematical point of view this means that the corresponding inverse problem might have no unique solution. To avoid such a situation we will restrict our consideration to cases where the refractive index along any optical trajectory does not vary noticeably, but an OPD exists between different (even adjacent) optical rays. In other words, we will consider the case when refractive index is a function of a single coordinate, n(y), and the gradients are not too large.

Figure 6.7 demonstrates the basic architecture for measuring temperature gradients. It is a dual-beam interferometer where a test object B is positioned in one branch and the reference branch either contains the same object, but at uniform temperature (B₁ shown by the dotted line), or has no object at all. A source S followed by lens L_1 provides monochromatic illumination for both channels. Lens L_2



FIGURE 6.7 (a) Basic configuration of an interferometer for measurement of temperature gradients and (b) schematic of the ray trajectories.

creates an image of the output surface of object B and reference B_1 onto the plane P where an interference pattern is generated.

The temperature gradients are supposed to be in the vertical direction. For example, a heater is positioned above sample B generating a one-dimensional heat flux Q so that the higher temperatures (and therefore the lower refractive index) are in the upper part of the object. As a light beam is traveling through the medium with variable refractive index the beam trajectory is no longer a straight line, but a trajectory with a curvature which depends on the derivative dn/dy (see Fig. 6.7b). To find the beam trajectory we begin with a differential equation describing the vertical coordinates of the beam, y(x), starting at height y_0 at x = 0 in the horizontal direction:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \left[1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\right] \frac{\mathrm{d}\ln n}{\mathrm{d}y} \tag{6.15}$$

and assume that the gradient of *n* is constant along the trajectory:

$$\frac{\mathrm{d}n}{\mathrm{d}y} = \mathrm{const} = K. \tag{6.16}$$

Then Eq. (6.15) is transformed to a new one

$$\frac{y''}{1+(y')^2} = \frac{1}{n_0} \frac{\mathrm{d}n}{\mathrm{d}y} = \frac{K}{n_0}$$
(6.17)

which has the exact solution

$$y(x) = -\frac{n_0}{K} \ln \cos\left(\frac{Kx}{n_0}\right) + y_0.$$
(6.18)

By expanding Eq. (6.18) in a series one obtains

$$y(x) = y_0 + \frac{Kx^2}{2n_0} \left(1 + \frac{K^2}{6n_0^2} x^2 + \dots + \right).$$
(6.19)

Hence, the first approximation is the parabolic function

$$y(x) - y_0 \approx \frac{Kx^2}{2n_0}.$$
 (6.20)

Once the beam trajectory function is known, we can find the optical paths, S_1 and S_2 , related to two separate starting points 1 and 2 along OY (see Fig. 6.7):

$$S_1 = \int_0^l n_1 \sqrt{1 + (y_1')^2} \, \mathrm{d}x; \quad S_2 = \int_0^l n_2 \sqrt{1 + (y_2')^2} \, \mathrm{d}x. \tag{6.21}$$

By calculating y'(x) either from Eq. (6.18) or from Eq. (6.20) and then substituting the obtained functions in Eq. (6.21) we can compute the OPD between these two paths.

Furthermore, taking into account

$$\frac{\mathrm{d}n}{\mathrm{d}y} = \frac{\mathrm{d}n}{\mathrm{d}T} \times \frac{\mathrm{d}T}{\mathrm{d}y} \tag{6.22}$$

where dn/dT is usually about $10^{-4} - 10^{-6}$ we draw the conclusion that if the temperature gradients are not very large a good approximation for the OPD is

$$\Delta_{21} = S_2 - S_1 = l(n_2 - n_1) = \Delta nl = \frac{\mathrm{d}n}{\mathrm{d}y} \times \Delta y \times l.$$
(6.23)

Now let optical paths 1 and 2 create two adjacent fringes in P. Then the OPD is equal to λ and the distance between fringes is expressed as follows:

$$\Delta y_{21} = \frac{\lambda}{l\frac{\mathrm{d}n}{\mathrm{d}y}} = \frac{\lambda}{l\frac{\mathrm{d}n}{\mathrm{d}T} \times \frac{\mathrm{d}T}{\mathrm{d}y}}.$$
(6.24)

This expression is the basis for converting the fringe spacing to temperature gradients. An example of a fringe pattern and the corresponding temperature profile is depicted in Fig. 6.8.

Problems

6.9. Find the accuracy of a parabolic approximation of the optical path in a layer of 70 mm in length with maximum temperature gradient $dT/dy = 20^{\circ}/\text{mm}$ if it is made of transparent material of n = 1.6; $dn/dT = 1 \times 10^{-4}$.



FIGURE 6.8 (a) Interference pattern and (b) corresponding temperature distribution.

6.10. In the optical configuration shown in Fig. 6.7a the imaging lens has a focal length of 100 mm and creates an interference pattern at magnification V = -1/3. Assuming the test object is of 50 mm in length and 15 mm in height and the maximum temperature gradient in it is 25°C/mm, find the minimum required diameter of the lens.

[Note: The refractive index of the object material is n = 1.3; and $dn/dT = 10^{-4}$].

6.11. The temperature profile in an object of 160 mm in length and 38 mm in height made of transparent material with n = 1.5 and $dn/dt = 10^{-6}$ is measured using the interferometric system shown in Fig. 6.7a. The interference pattern created at optical magnification V = -0.5 in the plane P includes 18 fringes with the following spacing: $d_1 = 0.12$ mm; $d_2 = 0.20$ mm; $d_3 = 0.27$ mm; $d_4 = 0.41$ mm; $d_5 = 0.63$ mm; $d_6 = 0.76$ mm; $d_7 = 0.94$ mm; $d_8 = 1.39$ mm; $d_9 = 1.56$ mm; $d_{10} = 1.80$ mm; $d_{11} = 1.82$ mm; $d_{12} = 1.88$ mm; $d_{13} = 1.80$ mm; $d_{14} = 1.74$ mm; $d_{15} = 1.74$ mm; $d_{16} = 1.04$ mm; $d_{17} = 0.48$ mm; $d_{18} = 0.48$ mm. Find the temperature profile in the sample if the minimum temperature (at the bottom) is 600°C.

6.4. Solutions to Problems

6.1. (a) The optical system is operated at magnification $V = S'/S \approx f'/S = -50/5,000 = -0.01$. Hence the area of the wall from which radiation reaches the detector is determined as

$$A = \frac{\pi (d_{\text{det}})^2}{4V^2} = \frac{9\pi}{4 \times 10^{-4}} = 0.0707 \text{ m}^2.$$

Due to the final size of the lens only a fraction of radiation emitted by *A* enters the detector and this fraction is

$$\frac{\omega}{2\pi} = \frac{\pi D^2}{2\pi \times 4 \times S^2} = \frac{30^2 \times 10^{-6}}{8 \times 5^2} = 4.5 \times 10^{-6}.$$

The total energy emitted by each 1 m² of the wall of T = 1,750 + 273 = 2,023 K in the spectral interval 0.4–1.1 µm can be calculated using the black body radiation table presented in Appendix 3:

$$E = \varepsilon \int_{0.4}^{1.1} e_{\mathrm{B}}(\lambda, T) \mathrm{d}\lambda = \varepsilon \left[\int_{0}^{1.1} e_{\mathrm{B}}(\lambda, T) \mathrm{d}\lambda - \int_{0}^{0.4} e_{\mathrm{B}}(\lambda, T) \mathrm{d}\lambda \right] = \varepsilon (I_2 - I_1)$$

where I_1 corresponds to $\lambda_1 T = 0.4 \times 2,023 = 809 \ \mu \text{mK}$ and I_2 corresponds to $\lambda_2 T = 1.1 \times 2,023 = 2,225 \ \mu \text{mK}$, so that interpolation between the

table data yields $I_2 - I_1 = \sigma T^4 \times 10^{-5} \times (1,057 - 0.172) \times 10^{-4} = 5.668 \times 10^{-4} (2,025)^4 1,056.8 \times 10^{-4} = 100,721 \text{ W/m}^2$. The power incident on the optical system is found as follows:

$$P = E \frac{\omega}{2\pi} A = 0.8 \times 100,721 \times 4.5 \times 10^{-6} \times 0.0707 = 2.562 \times 10^{-2} \text{ W}.$$

The detector maximum current of 1 mA originates from a power $P' = i_{det}R = 10^{-3}/0.28 = 3.54 \times 10^{-3}$ W. Therefore, the attenuator is required and the filter transmittance should be $\tau = 0.5 \times 3.54 \times 10^{-3}/2.562 \times 10^{-2} = 0.069$ (the factor of 0.5 takes into account that operation should be performed in the middle of the dynamic range).

(b) When the system is moved to the side the aperture angles of distant points will decrease whereas those of closer points will increase relative to the case of normal imaging, but on average we should not expect the geometry to change noticeably the detector reading. The same is true with regard to the projection of the wall segment imaged on the detector (projection of the radiated area on the direction perpendicular to the optical axis remains unchanged). However, the emissivity of the wall at 60° to the normal line for refractory materials is about 10% lower than the emissivity in the normal direction, and therefore the detector readings will be changed accordingly.

6.2. In temperature measurements at different angular positions the emissivity variation affects the amount of radiant energy registered by the detector. To estimate how much the emissivity can be changed we keep in mind Eq. (6.4) and calculate the reflectance of the smooth glass surface using Fresnel's formula (see Chapter 5). At the initial position (the incident angle $\varphi = 0^{\circ}$) we have $R_0 = (n-1)^2/(n+1)^2 = 0.04$; $\varepsilon_0 = 0.96$. At the first and the second angular positions the incident (and observation) angles are $\varphi_{\rm B} = \arctan(0.7) = 35.0^{\circ}$; $\varphi_{\rm C} = \arctan(1.4) = 54.5^{\circ}$. Calculating the corresponding refraction angles (from Snell's expression (1.2)) one obtains $r_{\rm B} = 22.48^{\circ}$; $r_{\rm C} = 32.9^{\circ}$ and proceeding further with Fresnel's formula we obtain $R_{\rm B} = 0.0429$ and $R_{\rm C} = 0.068$ which gives $\varepsilon_{\rm B} = 0.957$; $\varepsilon_{\rm C} = 0.932$. Now assuming that $\Delta T/T = \Delta \varepsilon/\varepsilon$ and keeping in mind that T = 773 K, we can conclude that the change of emissivity from ε_0 to $\varepsilon_{\rm C}$ results in $\Delta T_{\rm C} = 773 \times (3 \times 10^{-3}/0.96) = 23.0^{\circ}$.

6.3. Let us define the accuracy of Wien's formula (Eq. (6.9)) relative to the precise expression of Planck (Eq. (6.5)) by the ratio

$$q = \frac{e_{\rm B}^{(\rm W)}(\lambda, T)}{e_{\rm B}^{(\rm P)}(\lambda, T)} = \frac{\exp\left(-\frac{C_2}{\lambda T}\right)}{\left[\exp\left(\frac{C_2}{\lambda T}\right) - 1\right]^{-1}} = 1 - \exp\left(-\frac{C_2}{\lambda T}\right) = 1 - \Delta$$

which is obviously a monotonic function of a single parameter, λT : the greater this parameter the greater the relative error Δ . According to Wien's law (Eq. (6.6)) the wavelength of maximum emittance $\lambda_{max} = 2,898/T$ varies from 9.66 μ m to 0.966 μ m for the temperature range 300–3,000 K. In the relevant spectral interval ($\lambda_{max}/2$) < λ < $2\lambda_{max}$ the parameter λT varies from 1,449 to 5,796 for all temperatures and therefore $\Delta_{min} = \exp(-14.388/1,449) = 4.87 \times 10^{-5}$ and $\Delta_{max} = \exp(-14,388/5,796) = 0.0835$. Thus, the maximum error of Wien's formula in the chosen spectral and temperature intervals is about 8.4%.

6.4. Referring to Fig. 6.4 we consider the absorptance of the black body model for incident radiation E_{λ} directed at an angle φ . The incident beam illuminates a segment ds of the inner surface of the sphere which irradiates (reflects) radiation in all directions diffusively (uniformly). The fraction $(R \cos \varphi)E(\omega/2\pi)$ is lost due to reflection back to the entrance aperture (the factor $\cos \varphi$ takes into account that the normal to ds is not directed along the solid angle of emerging radiation) and the rest is absorbed completely inside the model. Therefore, the absorptance A_{λ} can be defined as follows:

$$A_{\lambda} = \frac{E_{\lambda} - R_{\lambda} \cos \varphi \times E_{\lambda} \frac{\omega}{2\pi}}{E_{\lambda}} = 1 - R_{\lambda} \cos \varphi \times \frac{\pi d^2 \cos \varphi}{4\rho^2} \times \frac{1}{2\pi}$$
$$= 1 - R_{\lambda} \cos \varphi \times \frac{\pi d^2 \cos \varphi}{4(D \cos \varphi)^2} \times \frac{1}{2\pi} = 1 - R_{\lambda} \left(\frac{d}{D}\right)^2 \times \frac{1}{8}.$$

For R = 0.8 and (d/D) = 0.05 we have $A_{\lambda} = 1 - 2.5 \times 10^{-4} = 0.99975$. This result is correct if we assume that the model is of uniform temperature and its inner surface is Lambertian (fully diffusive).

6.5. (a) From Eq. (6.12) we get

$$\frac{1}{T} = \frac{1}{2,200} - \frac{\ln(0.3/0.4)}{14,388 \times \left(\frac{1}{0.55} - \frac{1}{0.45}\right)}; \quad T = 2,468.8 \text{ K}.$$

(b) For the calculation of the brightness temperature at both wavelengths one can use Eq. (6.14):

For
$$\lambda = 0.45 \,\mu\text{m}$$
: $\frac{1}{T_{\text{BS1}}} = \frac{1}{T} - \frac{\ln(0.3)}{14,388} \times 0.45$; $T_{\text{BS1}} = 2258.8 \,\text{K}$
For $\lambda = 0.55 \,\mu\text{m}$: $\frac{1}{T_{\text{BS2}}} = \frac{1}{T} - \frac{\ln(0.4)}{14,388} \times 0.55$; $T_{\text{BS2}} = 2272.3 \,\text{K}$

Thus, the brightness temperature in both cases is significantly lower than the true temperature.

6.6. Referring to Fig. 6.6 we assume that both lenses L_1 and L_2 provide the same collection angle while imaging the test object M and the source S into the plane P observed by the eye through lens L_3 . Then the condition of equal brightness of both images in the plane P is expressed as

$$\varepsilon_{\lambda} e_{\rm B}(\lambda, T) = e_{\rm B}(\lambda, T_{\rm BS}) \times \tau_{\rm F1}$$
 (A)

where the transmittance of the filter F_1 is determined as follows (see also Section 5.1):

$$\tau_{F1} = (1 - R)^2 \exp(-Kt); \quad R = (n - 1)^2 / (n + 1)^2 = (0.6/2.6)^2 = 0.05325$$

$$\tau_{F1} = (1 - 0.05325)^2 \times \exp(-0.5 \times 2) = 0.3297.$$

By substituting this value in Eq. (A) and using Wien's function (Eq. (6.9)) we have

$$\frac{1}{T} - \frac{1}{T_{\rm BS}} = \frac{\ln(\varepsilon_{\lambda}/\tau_{\rm F1})}{C_2}\lambda; \quad \frac{1}{T} = \frac{1}{2,850} + \frac{\ln(0.5/0.3297)}{14,388}0.5; \quad T = 2737.1 \text{ K}.$$

6.7. Taking into consideration that the radiation temperature of the wall is 2,973 K and proceeding further with Eq. (6.10) we have

$$T = \frac{2,973}{\sqrt[4]{0.8}} = 3,143.6 \,\mathrm{K}.$$

Then for the heat flux emitted by the wall in the hemisphere we get

$$P = \varepsilon \int_{0.4}^{0.7} e_{\rm B}(\lambda, T) d\lambda = \varepsilon \left[\int_{0}^{0.7} e_{\rm B}(\lambda, T) d\lambda - \int_{0}^{0.4} e_{\rm B}(\lambda, T) d\lambda \right] = \varepsilon (I_2 - I_1)$$

and then calculating the integrals just as we did in Problem 6.1 we finally obtain $P = 0.8 \times 539.9 \times 10^3 = 432 \text{ kW/m}^2$.

6.8. The absolute color temperatures at the points of measurement are $(T_{\rm BC})_{\rm A} = 1,273$ K and $(T_{\rm BC})_{\rm B} = 923$ K. The ratio of the spectral emissivities required for Eq. (6.12) can be calculated as follows: $\varepsilon_{\lambda} = 1 - R_{\lambda} = \text{const}/\sqrt{\lambda}$; $\varepsilon_1/\varepsilon_2 = \sqrt{\lambda_2/\lambda_1} = \sqrt{0.52/0.45} = 1.075$. Then at each point the true temperature can be computed from Eq. (6.12) as

$$\frac{1}{T_{\rm A}} = \frac{1}{T_{\rm BC_A}} - \frac{\ln(1.075)}{14,388(1/0.52 - 1/0.45)}; \quad T_{\rm A} = 1,247.6 \text{ K}$$
$$\frac{1}{T_{\rm B}} = \frac{1}{T_{\rm BC_B}} - \frac{\ln(1.075)}{14,388(1/0.52 - 1/0.45)}; \quad T_{\rm B} = 909.3 \text{ K}$$

6.9. We characterize the accuracy of the approach by the error in finding the coordinates of the ray trajectory at the exit plane of the object, y(l). Assuming for simplicity $y_0 = 0$ we have for the parabolic approximation

$$K = \frac{\mathrm{d}n}{\mathrm{d}T} \times \frac{\mathrm{d}T}{\mathrm{d}y} = 10^{-4} \times 20 = 2 \times 10^{-3} \,^{\circ}\mathrm{C/mm};$$
$$y(l) = \frac{Kl^2}{2 \times 1.6} = \frac{2 \times 10^{-3} \times 70^2}{2 \times 1.6} = 3.063 \,\mathrm{mm}.$$

More accurate result can be obtained if the second term in the expansion (6.19) is taken into account. This term is $K^2 l^2 / 6n^2 = (4 \times 10^{-6} \times 4,900/6 \times 1.6^2) = 1.276 \times 10^{-3}$ mm which is obviously very small relatively to unity (the main term in the bracket on the right-hand side of Eq. (6.19)).

We can also estimate the error of the parabolic approximation by comparing the numerical result with that obtained from the precise solution (Eq. (6.18)). This accurate formula yields in our case

$$y(l) = -\frac{1.6}{2 \times 10^{-3}} \ln \cos\left(\frac{2 \times 10^{-3} \times 70}{1.6}\right) = 3.066 \text{ mm}$$

which differs from the parabolic approximation result by about 0.1%.

6.10. We suppose that the test sample is positioned symmetrically with regard to the optical axis of the interferometer imaging branch (see Fig. 6.7a). We also define the minimum diameter of lens L_2 as that which allows all rays leaving the exit of the test body to participate in creating the interference fringes in the plane P. Using the parabolic approximation (Eq. (6.20)) one can find the inclination angle of the rays at the exit from the sample:

$$y'(l) = \frac{2Kl}{2n} = \frac{10^{-4} \times 25 \times 50}{1.3} = 0.0962.$$

Since lens L₂ builds the image of the sample exit into the plane P at magnification V = -1/3, we can find the distance from the sample to the lens as follows (see Chapter 1):

$$S = f' \frac{1 - V}{V} = -100 \frac{1.333}{0.333} = -200 \text{ mm}$$

which enables one to determine the active size of the lens: $D_2 = 2[-y'(l) \times S + H/2] = 2 \times (200 \times 0.0962 + 7.5) = 53.4 \text{ mm}.$

6.11. To find the temperature profile from the interference pattern we rewrite Eq. (6.24) in the following manner:

$$\frac{\mathrm{d}T}{\mathrm{d}y} = \frac{\lambda}{l\frac{\mathrm{d}n}{\mathrm{d}T} \times \Delta y} \tag{A}$$

which enables one to find the temperature at the location of the *k*-th fringe if the temperature at location y_{k-1} is known:

$$T_k = T_{k-1} + \left(\frac{\mathrm{d}T}{\mathrm{d}y}\right)_k \times \Delta y_k; \quad (\Delta y_k = d_k).$$
 (B)

Starting from $T(0) = 600^{\circ}\text{C} = T_0$, we find $(dT/dy)_1$ from Eq. (A) with $\Delta y = d_1/V$, then calculate T_1 from Eq. (B) and proceed further until all fringes are interpreted. The numerical results are as follows:

$$\left(\frac{dT}{dy}\right)_{1} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 0.12} = 15.63^{\circ}\text{C/mm};$$

$$T_{1} = 600^{\circ} + 15.63 \times 0.12/0.5 = 603.8^{\circ}; \quad y_{1} = 0.24 \text{ mm}$$

$$\left(\frac{dT}{dy}\right)_{2} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 0.2} = 9.375^{\circ}\text{C/mm};$$

$$T_{2} = 603.8^{\circ} + 9.375 \times 0.20/0.5 = 607.6^{\circ}; \quad y_{2} = 0.64 \text{ mm}$$

$$\left(\frac{dT}{dy}\right)_{3} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 0.27} = 6.94^{\circ}\text{C/mm};$$

$$T_{3} = 607.6^{\circ} + 6.94 \times 0.27/0.5 = 611.3^{\circ}; \quad y_{3} = 1.2 \text{ mm}$$

$$\left(\frac{dT}{dy}\right)_{4} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 0.41} = 4.57^{\circ}\text{C/mm};$$

$$T_{4} = 611.3^{\circ} + 4.57 \times 0.41/0.5 = 615.0^{\circ}; \quad y_{4} = 2.02 \text{ mm}$$

$$\left(\frac{dT}{dy}\right)_{5} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 0.63} = 2.98^{\circ}\text{C/mm};$$

$$T_{5} = 615.0^{\circ} + 2.98 \times 0.63/0.5 = 618.8^{\circ}; \quad y_{5} = 3.28 \text{ mm}$$

$$\left(\frac{dT}{dy}\right)_{6} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 0.76} = 2.47^{\circ}\text{C/mm};$$

$$T_{6} = 618.8^{\circ} + 2.47 \times 0.76/0.5 = 622.6^{\circ}; \quad y_{6} = 4.80 \text{ mm}$$

$$\begin{pmatrix} \frac{dT}{dy} \end{pmatrix}_{7} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 0.94} = 2.00^{\circ} \text{C/mm}; \\ T_{7} = 622.6^{\circ} + 2.00 \times 0.94/0.5 = 626.4^{\circ}; y_{6} = 6.68 \text{ mm} \\ \begin{pmatrix} \frac{dT}{dy} \end{pmatrix}_{8} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 1.39} = 1.35^{\circ} \text{C/mm}; \\ T_{8} = 626.4^{\circ} + 1.35 \times 1.39/0.5 = 630.2^{\circ}; y_{8} = 9.46 \text{ mm} \\ \begin{pmatrix} \frac{dT}{dy} \end{pmatrix}_{9} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 1.56} = 1.2^{\circ} \text{C/mm}; \\ T_{9} = 630.2^{\circ} + 1.20 \times 1.56/0.5 = 633.9^{\circ}; y_{9} = 12.58 \text{ mm} \\ \begin{pmatrix} \frac{dT}{dy} \end{pmatrix}_{10} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 1.80} = 1.04^{\circ} \text{C/mm}; \\ T_{10} = 633.9^{\circ} + 1.04 \times 1.80/0.5 = 637.6^{\circ}; y_{10} = 16.18 \text{ mm} \\ \begin{pmatrix} \frac{dT}{dy} \end{pmatrix}_{11} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 1.82} = 1.03^{\circ} \text{C/mm}; \\ T_{11} = 637.6^{\circ} + 1.03 \times 1.82/0.5 = 641.4^{\circ}; y_{11} = 19.82 \text{ mm} \\ \begin{pmatrix} \frac{dT}{dy} \end{pmatrix}_{12} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 1.88} = 1.00^{\circ} \text{C/mm}; \\ T_{12} = 641.4^{\circ} + 1.88/0.5 = 645.0^{\circ}; y_{12} = 23.58 \text{ mm} \\ \begin{pmatrix} \frac{dT}{dy} \end{pmatrix}_{13} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 1.80} = 1.04^{\circ} \text{C/mm}; \\ T_{13} = 645.0^{\circ} + 1.04 \times 1.80/0.5 = 648.7^{\circ}; y_{13} = 27.18 \text{ mm} \\ \begin{pmatrix} \frac{dT}{dy} \end{pmatrix}_{14} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 1.74} = 1.08^{\circ} \text{C/mm}; \\ T_{14} = 648.7^{\circ} + 1.08 \times 1.74/0.5 = 652.5^{\circ}; y_{14} = 30.65 \text{ mm} \\ \begin{pmatrix} \frac{dT}{dy} \end{pmatrix}_{15} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 1.74} = 1.08^{\circ} \text{C/mm}; \\ T_{14} = 648.7^{\circ} + 1.08 \times 1.74/0.5 = 652.5^{\circ}; y_{15} = 34.12 \text{ mm} \end{pmatrix}$$

$$\left(\frac{dT}{dy}\right)_{16} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 1.04} = 1.80^{\circ}\text{C/mm};$$

$$T_{16} = 656.3^{\circ} + 1.80 \times 1.04/0.5 = 660.0^{\circ}; \quad y_{16} = 36.20 \text{ mm}$$

$$\left(\frac{dT}{dy}\right)_{17} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 0.48} = 3.91^{\circ}\text{C/mm};$$

$$T_{17} = 660.0^{\circ} + 3.91 \times 0.48/0.5 = 663.8^{\circ}; \quad y_{17} = 37.16 \text{ mm}$$

$$\left(\frac{dT}{dy}\right)_{18} = \frac{0.6 \times 10^{-3} \times 0.5}{160 \times 10^{-6} \times 0.47} = 3.99^{\circ}\text{C/mm};$$

$$T_{18} = 663.8^{\circ} + 3.99 \times 0.47/0.5 = 667.6^{\circ}; \quad y_{18} = 38.10 \text{ mm}.$$

Optical Scanners and Acousto-optics

7.1. Electro-mechanical Scanners

Optical scanners are instruments that cause a light beam to pass sequentially a number of spatial positions and to do so repeatedly, in a periodic manner. In most cases the scanning light is a laser beam.

There exists a great variety of opto-mechanical configurations capable of creating the scanning process. We will consider briefly four of them. The first is a simple plane mirror rotated around the axis normal to the plane of incident light (see Fig. 7.1). The scanning ray rotates in the plane of the figure and the angular velocity is twice the velocity of the rotating mirror: $\omega_S = 2\omega$ (if a mirror is turned by an angle φ the reflected ray is turned by 2φ). To convert the angular scanning



FIGURE 7.1 Single-mirror scanner: (a) incident and reflected rays; (b) location error along the scanning line.



FIGURE 7.2 Fast-rotating scanner: (a) basic configuration; (b) location of sources of error.

to parallel motion of the light beam a lens is sometimes added after the mirror, providing the focus of the lens coincides with the rotation axis.

The optimal operation of the mirror scanner requires that the axis of rotation be in the reflection plane of the mirror, otherwise an error in the beam displacement along the scanning line might occur. The explanation of this error is demonstrated in Fig. 7.1b (segment Δ) and details of its calculation are presented in Problem 7.1. The other error of this simple scanner is related to the variation of linear velocity of the light spot along the scanning line. Due to this error the time interval in which a point on the scanning line is exposed to radiation is not equal for different locations along the line (exposure error, ΔE_{exp}). Details of this error are discussed in Problem 7.2.

A single mirror is also exploited in fast-rotating scanners like that shown in Fig. 7.2. In this case the rotation axis and the incident light beam coincide with each other and the mirror surface is tilted at 45° to both of them. The scanning beam moves (rotates) in the plane normal to the rotation axis. However, because of misalignment errors the scanning plane might be tilted or even transformed into a conic surface. The three main sources of errors are pointed out in Fig. 7.2b. As can be understood from a simple geometrical consideration, displacement and tilt misalignments cause tilting of the scanning plane whereas an error in the 45° angle changes the shape of the scanning surface. Numerical results can be found in Problem 7.3.

Another architecture of a fast scanner is a polygon (a mirror drum) where a number of mirrors are rotated together around a single axis. The example depicted in Fig. 7.3 shows the creation of a two-dimensional (2-D) raster on a sheet of paper or film. Scanning in the OX direction is performed by the mirror drum while motion in the OY direction results from the mechanical motion of the sheet itself. Transfer from one mirror of the drum to another constitutes the sequential horizontal lines of the raster. The most significant error is "wobbling" of the raster lines: since the mirror surfaces are not exactly parallel to the rotation axis, each mirror reflects the incident beam in a slightly different way. As a result, the spacing of the raster



FIGURE 7.3 (a) Polygon (mirror drum) scanner and (b) 2-D raster with spacing error Δ_S .

lines is not constant, but varies according to the wobble angles of the mirrors (see Fig. 7.3b). Additional tilt of some raster lines might also occur. The number of mirrors together with the distance to the scanning surface dictate the length of the raster lines: the greater this number the faster the scanning process, but the shorter the raster lines (see the calculation in Problem 7.4).

A galvanometer scanner converts directly input electric signals into the angular position of a scanner element – a small mirror M coupled to a moving coil or to a solid iron rotor (magnetic driver, see Fig. 7.4). Compared to other mirrorbased scanners the galvanometer device enables one to address any location of the scanning path independent of the previous position of the light beam. This great advantage allows the creation of complex trajectories (latent "picture") on a scanning surface exposed to radiation in any predetermined manner (like characters, maps, etc.). Although the mirror is small and light, mechanical inertia is still a problem and standard scanners perform a sawtooth of the raster or random stepping traveling at a frequency of up to 1 kHz. Faster operation is achieved in resonant scanners which produce a sinusoidal scanning motion at frequencies of about 5–10 kHz. Since the mirror motion (oscillation) is limited



FIGURE 7.4 Galvanometer scanner.



FIGURE 7.5 Basic configuration of a 2-D galvanometer scanner.

by the maximum excursion angle of about 30–60°, the angular velocity is not constant, but varies along the oscillation trajectory. This results in the exposure error, ΔE_{exp} , mentioned above. The wobble caused by random fluctuations of the moving electro-mechanical parts is kept at a level of 1–10 arcseconds.

Due to their relatively compact design galvanometer scanners are commonly used in pairs in systems intended for 2-D scanning procedures. Such a configuration is depicted in Fig. 7.5.

Problems

7.1. A single-mirror scanner located 1 m from a scanning surface provides a scanning line of 80 cm in length. Assuming the incident light angle in the zero (reference) position is 30° and the rotation axis is 2 cm behind the mirror surface, calculate the maximum location error along the scanning line.

7.2. In a single-mirror scanner the rotation speed is 120 rpm, the excursion angle $2\varphi_{\text{max}} = 60^{\circ}$, and the distance to the scanning surface $L_0 = 1$ m. Find the exposure and the maximum exposure error along the scanning line if the scanner is operated with a laser beam of 5 mW total power, 0.5 mm waist, and divergence angle: (a) $2\theta = 2 \times 10^{-5}$; (b) $2\theta = 4 \times 10^{-3}$.

7.3. In a fast-rotating scanner followed by a lens (60 mm active diameter, 120 mm focal length) and working with a laser beam of very small cross-section, the following misalignment is revealed: (a) parallel displacement of 0.5 mm; (b) tilt of 0.5° . Find the location errors in both cases and determine which misalignment is more critical.

7.4. A mirror drum scanner (a polygon scanner) with six parallel mirrors provides raster scanning on a paper sheet of 1 m in length.

- (a) Find position of the scanning area relative to the drum if a laser beam incident on the mirrors in its medium position is parallel to the paper surface.
- (b) Calculate the raster errors if the wobble angles of the mirrors are in the region of 2 arcseconds.

7.2. Acousto-optics and Acousto-optical Scanners

7.2.1. Acousto-optical Effect and Acousto-optical Cell (AOM)

The subject of acousto-optics is concerned with the interaction of light with acoustic waves. If electromagnetic waves propagate through a medium where acoustic waves are generated the spatial distribution of light becomes noticeably dependent on the parameters of the acoustic waves. This phenomenon, frequently termed light scattering on acoustic waves or simply as the acousto-optical effect, is widely exploited in numerous optical systems including optical scanners.

The physical basis of the phenomenon is the fact that acoustic waves which are actually periodic variations of the density of the medium cause corresponding periodic variations of the refractive index affecting the propagation of electromagnetic waves. The relation between refractive index, n, and density, ρ , is governed by the Gladston–Dale formula which for the simple case of gases has the form

$$\frac{n-1}{\rho} = K \tag{7.1}$$

resulting from the well-known and more general Lorentz–Lorenz formula (e.g., see Born and Wolf, 1968)

$$\frac{n^2 - 1}{n^2 + 2} \times \frac{1}{\rho} = A/W = \text{const}$$
 (7.2)

where A is the molar refraction and W is the molecular weight of the substance.

As far as dynamic phenomena (like wave propagation) are concerned variations of refractive index are related to the mechanical strain components in a material, S_i (a detailed description can be found in Yariv, 1984):

$$\Delta\left(\frac{1}{n}\right)_{i} = p_{ij}S_{j}; \quad (i, j = 1, 2, 3)$$
(7.3)

where material properties are described by the tensor p_{ij} . It is quite understandable that in some materials the refractive index varies noticeably in response to mechanical stresses and in some others it does not. Also, in the same material the reaction to shear stresses could be much greater than to longitudinal ones, so that the usefulness of the material for the generation of the acousto-optical effect depends on how the sample is prepared and operated.

Of the numerous parameters characterizing acoustic properties of a material the most important for acousto-optics is acoustic velocity, V_s . This varies from 620 m/s for TeO₂ (for shear stresses mode) to 5,000–6,000 m/s for LiNbO₃ and quartz. The combination of all relevant properties of acousto-optical materials are arranged in a single figure of merit, M, and this parameter solely characterizes the acousto-optical behavior of a material as far as the efficiency of acousto-optical cells is concerned.

In general, an acousto-optical cell (more frequently termed an acousto-optical modulator, AOM, and sometimes also called a Bragg cell) is a slab of optically transparent material coupled acoustically (mechanically) to a transducer T (usually piezoelectric) which converts incoming electrical signals of high frequency into acoustic oscillations propagating in the slab. The cell is illuminated by a parallel light beam, in most cases – from a laser source. The basic arrangement of an AOM is shown in Fig. 7.6. A monochromatic beam of wavelength λ is incident at an angle θ on the slab from the left and an acoustic wave of wavelength Λ propagates in the direction OY. This wave is generated by a transducer T fed by an RF signal of frequency *f* originating in an external driver: $\Lambda = V_S/f$. Usually *f* is in the range 30–1,000 MHz, and Λ is from several micrometers to several tens of micrometers,



FIGURE 7.6 Basic configuration of (a) an AOM and (b) diagrams of up-shifted and down-shifted beams.

depending on the material used. On the upper (opposite to T) side of the cell special measures are taken (like acoustic absorbers, etc.) in order to reduce the reflection of the acoustic waves back to the slab, since AOMs are mostly operated with free traveling waves (not with standing waves, although this is also theoretically possible).

The diameter of the light beam, D, has to be significantly greater than the acoustic wave period, Λ . Then, for a beam traveling through the AOM the cell operates like a phase diffraction grating (see explanation of diffraction gratings in Chapter 5). This means that not the amplitude (transparency) but the phase of the propagating beam is changed periodically, according to the variation of the optical path, nL, while moving across the beam. As a result, the light at the output is unequally distributed in space, with some directions of strong maxima followed by intervals of negligible intensity. These maxima are called the first, second, etc., diffraction orders, the zero order being the unshifted incident beam. There also exist beams of (-1)st, (-2)nd, etc., diffraction orders on the other side of I_0 . It is the diffraction orders (usually only the first one or two) that are exploited in AOM applications, and the higher the intensity of the diffraction orders the higher the efficiency of the cell. It can be shown that the best condition (the highest efficiency) is achieved if the incidence angle obeys the Bragg condition (Bragg incidence angle):

$$\theta = \theta_{\rm B} = \frac{\lambda}{2 \times \Lambda} = \frac{\lambda}{2V_{\rm S}}f.$$
(7.4)

There are two possible ways to obey this condition, with the up-shifted and downshifted beams, as shown in Fig. 7.6b where the wave vectors $\vec{K_0}, \vec{K_1}, \vec{K_{-1}}$, and $\vec{K_S} \left(\left| \vec{K} \right| = 2\pi/\lambda; \left| \vec{K_S} \right| = 2\pi/\Lambda \right)$ are depicted. As we see, in both cases the firstorder diffraction beam is separated by the angle $2\theta_B$ from the zero order. It should also be noted that the wavelengths in the diffraction orders are different from each other and from the zero order, so that there is no way to get an interference pattern if these beams are overlapped anywhere after leaving the AOM.

The efficiency of an AOM calculated for the first diffraction order is governed by

$$\eta = \frac{I_1}{I_0} = \sin^2\left(\frac{\pi}{2}\sqrt{\frac{2LMP_{\rm ac}}{\lambda^2 H}}\right) \tag{7.5}$$

where *M* is the figure of merit of the AOM material, *H* is the height of the cell (in the direction perpendicular to the plane of Fig. 7.6), and P_{ac} is the acoustic power transferred by the transducer to the cell. In most cases the value under the square root is small enough so that the sine term can be replaced just by the argument. This leads to a linear relation between the applied acoustic power and the AOM efficiency, i.e., the intensity of light in the first diffraction order becomes proportional to P_{ac} .

Problems

7.5. A laser beam of 0.5 μ m wavelength strikes a Bragg cell made of LiNbO₃ ($V_S = 3,400$ m/s) and energized by an RF signal of 100 MHz. Find the angular separation between the first order and the (-1)st order diffracted beams.

7.6. What is the maximum rate of input signal variation if it is processed by an acousto-optical system with a Bragg cell of 15 mm in length made of TeO₂ (shear mode, $V_S = 620$ m/s) illuminated by a laser beam of 5 mm in diameter?

7.7. Dual-path arrangement. For improving the contrast of the diffracted beam with regard to the background light, a double-path arrangement has been suggested where the laser beam travels twice through a Bragg cell crystal. Assuming an AOM of 3 mm (H) × 6 mm (L) size made of PbMoO₄ ($V_S = 3,400 \text{ m/s}$; $M = 10^{-6}$) is energized by acoustic power $P_{ac} = 0.03$ W and illuminated by a laser beam of 10 mW power, 1 mm diameter, and 0.6 µm wavelength, find the optimal configuration of the system and calculate the light intensity of the first order diffracted beam.

7.8. Spectral imaging. A transparent object P illuminated by a white light source is imaged into a plane M by two identical lenses L_1 and L_2 , as shown in Fig. 7.7. When an AOM (made of TeO₂; $V_S = 620$ m/s) located between the lenses is energized by an RF signal of 80 MHz the images of different wavelengths are angularly separated and only those which correspond to wavelengths of maximum transparency of filters F_1 and F_2 reach the area sensors CCD₁ and CCD₂ (2.4 mm × 1.8 mm each). In such a way images from a chosen pair of wavelengths, λ_1 and λ_2 , are compared in real time. Choosing new wavelengths is accompanied by moving the CCDs and the filters to a new (optimal) position and changing the RF frequency.

(a) Assuming the minimum spacing between the CCDs to be 2 mm and the chosen wavelengths are $\lambda_1 = 0.65 \ \mu$ m, and $\lambda_2 = 0.55 \ \mu$ m, find the minimum



FIGURE 7.7 Problem 7.8 – Optical system for spectral imaging.

possible focal length of the lenses and the location of the sensors in the plane M. [Note: Imaging is carried out in the first diffraction order.]

(b) How different should the output signals in both channels be expected for the same radiation level in the plane P if the CCDs are made of silicon having a spectral quantum efficiency of 32% and 21% for the two wavelengths, respectively?

7.2.2. Two Operation Modes: AOM as Modulator of Light and AOM as Deflector of Optical Beams

Depending on the geometry of the Bragg cell and the angle of the incidence beam an AOM can be operated in two different modes: as a modulator or as a deflector. In the first case the intensity of light (usually of the first diffracted beam) is modulated by changing the input acoustic power, with no change of the RF signal frequency and consequently with no change of the spatial location of the receiver (detector assembly). Sometimes information is transferred simultaneously to several receivers (e.g., one is illuminated by the first diffraction order and another by the (-1)st diffraction order), but again the position of the receivers in space remains constant. Examples of Bragg cells operated in the modulation mode are considered in Problems 7.5–7.8. It should be noted that varying the light intensity can obviously be done by changing the optical power: if a laser diode or even a LED are used as the light source, the optical power is easily controlled by changing the electric current supplied to the source. However, modulation by acoustic power has some significant advantages which become crucial in a number of applications (an important example is considered in Section 7.2.3).

When a Bragg cell is used as a deflector the carrier of the acoustic wave supplied to the AOM is changed, usually in some specific manner, like a sawtooth for line scanners, for instance. Then the direction of the diffracted beam is varied accordingly and the light beam travels along the scanning axis, with no involvement of mechanical or electro-mechanical moving parts, as it was in the cases described in Section 7.1.

The variation of RF frequency, Δf , is related to the change of the diffraction angle, $\Delta \theta$, as

$$\Delta \theta = \Delta f \times (\lambda/V_{\rm S}). \tag{7.6}$$

This does not mean, however, that any desirable spatial position can be precisely achieved. There is a basic limitation of the angular spatial resolution caused by diffraction. That is, if a light beam of size D is propagated through the AOM, diffraction not only rearranges the light into diffraction orders, but also changes the specific shape of each diffracted beam, so that the main diffraction spot has



FIGURE 7.8 Deflection of light by an AOM.

an angular width of $\delta\theta = \lambda/D$ and the radial intensity distribution is described by the Airy (or Gaussian-like) function (see details in Chapters 2 and 5). Therefore, the whole range of angular variation is subdivided into spots of finite width (see Fig. 7.8). The number of resolvable spots, N, can be determined in terms of a time-bandwidth product, TBW = $\tau \times \Delta f$, if we take into account that the time needed for an acoustic wave to pass the light beam is $\tau = D/V_S$ and substitute this expression in Eq. (7.6):

$$N = \frac{\Delta\theta}{\delta\theta} = \frac{(\lambda/V_{\rm S}) \times \Delta f}{\lambda/D} = \tau \times \Delta f.$$
(7.7)

Usually for a good-quality system TBW is of the order of 10^4 .

A line scanner with an AOM in the deflection mode is presented in Fig. 7.9. A laser beam moves in the OX direction and is controlled by a signal generated in an RF driver. The time history of a typical signal shows a frequency variation in



FIGURE 7.9 (a) Line scanner with AOM and (b) time history of the RF signals.

the range from f_{\min} to f_{\max} which is dictated by the properties of the Bragg cell crystal. It is obvious that while moving from A to B the light can also be modulated with regard to its amplitude or even switched on and off according to a program prepared in advance and enabling one to expose the paper or film to radiation in any desirable manner and at very high speed.

Problems

7.9. An optical communication system comprises an AOM and Nd:YAG laser (wavelength 1.064 μ m, beam diameter 1 mm). The AOM is made of TeO₂ ($V_S = 620$ m/s) and operates around a 50 MHz RF signal. Information is transferred to two receivers located 100 m from one another each at a distance of 2 km from the transmitter.

- (a) Find the range of RF signals required for communication.
- (b) How many receivers can be simultaneously treated by the system without cross-talk between them?

7.10. Calculate the number of resolvable angular locations of an AOM deflector operated with a PbMoO₄ crystal ($V_S = 4,200$ m/s) and illuminated by a He–Ne laser beam expanded to 5 mm in diameter if the crystal acoustic efficiency varies with frequency as shown in Fig. 7.10.

7.2.3. AOM Architecture for Spectral Analysis

Acousto-optical architecture can be exploited for the spectral analysis of fast electrical signals. An example of such a system is shown in Fig. 7.11. A laser beam is expanded by a standard beam expander and strikes an AOM at the Bragg angle corresponding to a frequency f_C of the carrier acoustic wave. An RF driver comprises a carrier oscillator and a mixer which supplies an RF frequency modulated by a test signal S(t) to the AOM.



FIGURE 7.10 Problem 7.10 – Bragg cell efficiency vs. RF frequency.


FIGURE 7.11 AOM architecture for spectral analysis.

Propagation of the acoustic waves of modulated amplitude is equivalent to propagation of a collection of sinusoidal signals simultaneously through the Bragg cell and each harmonic generates a separate diffracted beam the intensity of which is dictated by the harmonic amplitude. The procedure is equivalent to expanding S(t) into Fourier components, each component being focused by the collection lens in its focal plane where the detector array (photodiodes or CCD) is located.

More specifically, let the carrier oscillation be $A \cos(2\pi f_C t)$ and the test signal be a single harmonic of frequency $f: S(t) = mA \cos(2\pi ft)$. Then, the signal after the mixer is $A[1+S(t)] \times \cos(2\pi f_C t)$ and the corresponding acoustic wave U(x, t) in the Bragg cell becomes

$$U(x,t) = C_0 \cos\left[2\pi f_{\rm C}\left(t - \frac{x}{V_{\rm S}}\right)\right] + C_1 \cos\left[2\pi (f_{\rm C} + f)\left(t - \frac{x}{V_{\rm S}}\right)\right] + C_2 \cos\left[2\pi (f_{\rm C} - f)\left(t - \frac{x}{V_{\rm S}}\right)\right].$$
 (7.8)

Thus, a single harmonic acts as three waves propagating simultaneously through the AOM: the first is of the carrier frequency f_C and it corresponds to the center of the first diffraction order; the second is of frequency $(f_C + f)$ and it is tilted by an angle $\theta' = (\lambda/V_S)f$ from the center of the first diffraction order; and the third of frequency $(f_C - f)$ and it is tilted by an angle $(-\theta')$ from the direction of the first diffraction order. Therefore, a single harmonic results in three light spots in the detector plane (the focal plane of the lens), and this is also true for any other Fourier component of the signal S(t). Actually to reveal the presence of any frequency f in the test signal it is enough to analyze half of the first diffraction interval (either up-shifted or down-shifted). The length of the detector array should be defined accordingly.

The spectral resolution (the minimum resolvable frequency interval, δf) depends on the minimum size of the light spot in the detector plane and it evidently depends on the diffraction limit of the system. The number of resolvable spectral components, $N = \Delta f/\delta f$, is defined in the usual manner (see Eq. (7.7)). Obviously the pitch of the detector array (or the size of a single element) has to be compatible with the spectral resolution (diffraction spot size).

It is worth pointing out that we have addressed here the spectral analysis of electrical signals only. The spectral analysis of optical signals is discussed in detail in Chapter 5.

Problems

7.11. An optical system like that of Fig. 7.11 is built around an AOM made of TeO₂ ($V_S = 620$ m/s, f = 40-80 MHz) and operated with a Ga–As laser diode (wavelength of 0.83 µm) followed by an anamorphic collimator providing a light beam of elliptical shape with maximum size of 10 mm.

- (a) Find the spectral resolution of the system if the Bragg cell length is 20 mm.
- (b) The Ga–As laser is replaced by another one which generates a beam of 0.6 μm wavelength. How does this affect the system performance?

7.12. A resolution of 30 kHz is required from an acousto-optical system for spectral measurement. An available detector array is a line CCD of 1,024 elements, 0.015 mm \times 2 mm each. How should the rest of the system be configured?

7.3. Solutions to Problems

7.1. Referring to Fig. 7.12, we assume that the scanning line AB = 80 cm is normal to the reflected ray OC corresponding to the zero position of the mirror. Hence, AC = BC and the mirror is rotated in the range of $\pm \varphi_{max} = (1/2) \arctan(BC/OC) = 0.5 \arctan 0.4 = 10.9^{\circ}$. Furthermore, because the point of light incidence moves along the mirror surface from O to M the reflected beam also moves from position OB to its real position MN, so that the location error on the scanning line is $\Delta_l = BN$. To calculate BN we first find the segment Δ_0 perpendicular to the reflected beam:

$$\Delta_0 = OM \sin[180 - 2(i + \varphi)] = OM \times \sin[2(i + \varphi)].$$
(A)



FIGURE 7.12 Problem 7.1 – Geometry of location error.

Taking into account that in triangle QOM the side OQ is equal to $z = (R/\cos\varphi) - R$ and therefore

$$OM = z \frac{\sin(180^\circ - 90^\circ + \varphi)}{\sin[180^\circ - i - (180^\circ - 90^\circ + \varphi)]} = z \frac{\cos\varphi}{\cos(\varphi + i)} = R \frac{(1 - \cos\varphi)}{\cos(\varphi + i)}$$

we have from Eq. (A):

$$\Delta_0 = R \frac{1 - \cos \varphi}{\cos(\varphi + i)} \times \sin[2(\varphi + i)] = 2R(1 - \cos \varphi) \times \sin(\varphi + i)$$
(B)

and finally the location error $\Delta_l = \Delta_0/\sin \alpha = \Delta_0/\cos(2\varphi)$. By substituting in Eq. (B) the data of the problem and the found value of φ_{max} , we get $\Delta_0 = 2 \times 20 \times (1 - \cos 10.9^\circ) \times \sin(30^\circ + 10.9^\circ) = 0.47 \text{ mm}; \Delta_l = 0.47/\cos(21.8^\circ) = 0.506 \text{ mm}.$

7.2. The geometry of scanning is depicted schematically in Fig. 7.13. We define the exposure E_{exp} of an element $\Delta x \times \Delta y$ by the following integral:

$$E_{\exp} = \Delta x \times \Delta y \times \int_{0}^{\tau} I[x(t), y] dt$$
 (A)

where τ is the exposure time (the time interval in which the point of interest, M(*x*), is exposed to the radiation of the scanning beam traveling through M). We also suppose that the angular velocity of the rotation, ω , is constant (no noise or random



FIGURE 7.13 Problem 7.2 – Scanning beam at two angular positions.

oscillations) so that the linear velocity, V, is determined as follows:

$$x = L \times \tan \varphi; \quad V = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{L}{\cos^2 \varphi} \frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{L\omega}{\cos^2 \varphi}.$$
 (B)

For simplicity we will consider first the situation when the scanning beam is of constant light intensity, I_0 , across the beam diameter $D_0 = 2w_0$ as well as at different distances L from the mirror. Then, keeping in mind that $I_0 = 4P/\pi D_0^2$ and $\tau_0 = D_0/V_0$, we get from Eq. (A) the exposure at the initial position of zero angle ($\varphi = 0$):

$$E_{\exp 0} = \Delta x \times \Delta y \times I_0 \tau_0 = \Delta x \times \Delta y \times \frac{4P}{\pi D_0} \times \frac{1}{\omega L}$$
$$= \Delta x \times \Delta y \times \frac{4 \times 5 \times 10^{-3}}{\pi \times 1 \times 2 \times 10^3} = 3.184 \,\mu\text{J/element.}$$

To calculate the exposure for angular position φ we should remember that the light intensity incident on the element $\Delta x \times \Delta y$ is $I_0 \cos \varphi$ and the length of the light spot traveling through point M is $D_0/\cos \varphi$. Then, using Eq. (B) we obtain

$$E_{\exp\varphi} = \Delta x \times \Delta y \times I_{\varphi} \tau_{\varphi} = \Delta x \times \Delta y \times \frac{4P \cos\varphi}{\pi D_0^2} \times \frac{D_0/\cos\varphi}{\omega L/\cos^2\varphi} = E_{\exp 0} \cos^2\varphi$$

and the maximum exposure error is

$$\Delta_{\exp} = E_{\exp 0} - E_{\exp \varphi} = E_{\exp 0} (1 - \cos^2 \varphi_{\max}) = 3.184(1 - 0.75)$$

= 0.8 \mu J/element

that is, about 25%.

Returning to the general case we consider a laser beam with a Gaussian light distribution, as per Eq. (3.2), with intensity in the middle point defined by Eq. (3.6) and the beam waist size as in Eq. (3.1) (see also Eq. (B) of Problem 3.7). We also assume that the light spot is symmetrical, so that the exposure in Eq. (A) can be calculated as twice the integral from 0 to $\tau_{exp}/2$. Expression (A) yields

$$E_{\exp} = \Delta x \times \Delta y \times \frac{2P}{\pi w_{\rm L}^2} \cos \varphi \times \frac{w_{\rm L}}{\sqrt{2}V \cos \varphi} \times 2 \int_0^{\tau_{\exp}/2} \exp\left(-\frac{2V^2 \cos^2 \varphi}{w_{\rm L}^2}t^2\right) dt$$
$$= \Delta x \times \Delta y \times \cos \varphi \frac{P}{w_{\rm L}} \times \frac{\sqrt{2/\pi}}{V \cos \varphi} \operatorname{erf}\left(\frac{V \cos \varphi}{w_{\rm L}\sqrt{2}}\tau_{\exp}\right)$$
$$\approx \Delta x \times \Delta y \times \frac{\sqrt{2/\pi}P}{\omega L \times w_{\rm L}} \cos^2 \varphi \tag{C}$$

where the error function is taken as unity because of the relatively large value of its argument. Proceeding to cases (a) and (b) of the problem, we should remember that $w_L^2 = w_0^2 + (\theta L/\cos\varphi)^2$. (a) $\theta = 10^{-5}$:

$$w_{\rm L} = \sqrt{0.25 + (10^{-4}/\cos^2 \varphi)} \approx 0.5;$$

$$E_{\rm exp} = {\rm const} \frac{\sqrt{2/\pi} \times 5 \times 10^{-3}}{2 \times 10^3 \times 0.5} \cos^2 \varphi = E_0 \cos^2 \varphi$$

Hence, the exposure error varies as $1 - \cos^2 \varphi$ increasing again to about 25% for $\varphi_{\text{max}} = 30^\circ$. (b) $\theta = 2 \times 10^{-3}$:

$$w_{\rm L} = \sqrt{0.25 + (4/\cos^2 \varphi)};$$
 $w_{\rm L}(0) = 2.062 \text{ mm};$ $w_{\rm L}(30^\circ) = 2.36 \text{ mm};$
 $\Delta E_{\rm exp}/E_{\rm exp}(0) = 1 - \frac{\cos^2(30^\circ) \times 2.06}{2.36} = 0.345 \text{ or about } 35\%.$

7.3. The misalignment errors for both cases are demonstrated in Fig. 7.14.

(a) As shown in Fig. 7.14a, the parallel displacement of the incident laser beam causes the reflected beam to cross the lens along the curve (dotted line shown in the figure). This results in both a horizontal displacement $2\Delta_x = 2\Delta \sin \alpha$ (where $\tan \alpha = D/2f' = 60/240 = 0.25$; $\alpha = 14^\circ$ therefore $2\Delta_x = 0.24$ mm) and a tilt of $\psi = \arctan(\Delta/f') = \arctan(0.5/120) = 4.17 \times 10^{-3}$.

(b) The tilt misalignment of 0.5° causes inclination of the scanning plane, so that the reflected beam will cross the lens along the curve shown in Fig. 7.14b. Since in any case the beam striking the lens comes from its focal point, there is no tilting



FIGURE 7.14 Problem 7.3 – Location errors of a fast-rotating scanner originating from (a) displacement and (b) tilt of the incoming laser beam.

behind the lens (at any position the beam is parallel to the lens axis). However, parallel displacement will occur, reaching the value $\Delta_h = \theta f' = 30 \times 3 \times 10^{-4} \times 120 = 1.08$ mm in the horizontal plane and increasing further while moving to point A on the periphery of the lens (for more accurate results a consideration of the geometry of a cone crossing the lens should be carefully carried out).

Therefore, the misalignment tilt yields a location error which is more than four times greater than the error caused by the parallel displacement.

7.4. (a) Referring to Fig. 7.15, we see that the polygon rotation by an angle $\varphi = 360^{\circ}/N = 60^{\circ}$ (*N* is the number of mirrors) results in beam travel over the whole scanning length AB. This obviously yields the distance $L = H/2 \tan \varphi = 1/2\sqrt{3} = 0.29$ m. As the incident beam should be parallel to AB and the reflected beam at the middle position is perpendicular to AB, we get the angle of incidence $i = 45^{\circ}$.

(b) A wobbling angle θ causes a raster spacing error $\Delta_S = L\theta = 0.29 \times 2 \times 5 \times 10^{-6} = 2.9 \,\mu\text{m}$. The raster tilt error is equal to zero in this case because all raster lines remain parallel to the incident beam which is parallel to the paper surface.

7.5. The optimal angle of light incidence on the AOM is calculated from the Bragg condition (Eq. (7.4)):

$$\theta_{\rm B} = \frac{\lambda}{2V_{\rm s}} f = \frac{0.5 \times 10^{-6}}{2 \times 3,400} 100 \times 10^{6} = 0.0073.$$



FIGURE 7.15 Problem 7.4 – Geometry of a six-mirror scanner.

This angle determines the direction of the zero order (unshifted beam). Then, the first diffraction order is shifted up from the zero order by $2\theta_B$ and the (-1)st diffraction order is shifted down by the same angle, yielding an angular separation $4\theta_B = 0.0292 = 1.673^\circ$ between these two diffraction beams. As we see, the angles are very small, meaning that precise alignment of the AOM and related optics is required.

7.6. Since the length of the AOM is larger than the light beam size, the limitation in signal variation rate originates from a minimum time interval required for the acoustic waves to pass through the laser beam: $\tau_{\rm min} = D/V_{\rm S} = (5 \times 10^{-3}/620) = 8.06 \,\mu s$. Therefore, the maximum frequency (the maximum rate) is defined by reciprocal of twice of minimum time interval:

$$f_{\text{max}} = \frac{1}{2\tau_{\text{min}}} = \frac{10^6}{2 \times 8.06} = 62 \text{ kHz}.$$

7.7. A possible dual-path arrangement allowing for an increase of the contrast ratio is shown schematically in Fig. 7.16. The initial horizontal light beam strikes an AOM at the Bragg angle θ_B and the first order diffracted beam is defined by the vector $\vec{K_1}$ tilted at $2\theta_B$ to the horizontal axis. This beam is incident on a retroreflector, R, and is then reflected back to the AOM at the same angle. Hence, the vector $(-\vec{K_1})$ strikes the AOM for the second time, again at the Bragg angle. As a consequence, the new up-shifted beam (the first diffraction order of the inverted first-order beam) is defined by the vector $-\vec{K_{11}}$ which is parallel to the initial beam. This latter beam (reflected aside from the arrangement by a mirror M, for instance) can be used in further applications. Diaphragms can be used in order to prevent the



FIGURE 7.16 Problem 7.7 – Dual-path arrangement with an AOM.

background light originating in the zero-order beam of each path from propagating in the outgoing direction and hence to improve the contrast ratio.

To calculate the intensity of the outgoing beam we first find the intensity of the initial beam using Eq. (3.6) and keeping in mind that w = D/2 = 0.5 mm:

$$I_0 = \frac{2P}{\pi w^2} = \frac{2 \times 10}{\pi \times 0.25} = 25.47 \text{ mW/mm}^2.$$

Then we proceed to the first-order beam in the single path and use the definition of AOM efficiency (Eq. (7.5)):

$$\eta = \sin^2 \left(\frac{\pi}{2} \sqrt{\frac{2 \times 6 \times 10^{-6} \times 0.03}{0.6^2 \times 10^{-6} \times 3}} \right) = \sin^2(0.907) = 0.62;$$
$$I_1 = I_0 \eta = 15.8 \text{ mW/mm}^2.$$

Finally for the second path we have $I_{11} = I_1 \eta = 9.80 \text{ mW/mm}^2$.

In these calculations we take into account that the argument of the sine term in Eq. (7.5) is not small enough for a linear approximation (the device is operated in the non-linear range with regard to the acoustic power).

7.8. (a) Since the AOM is operated simultaneously at two wavelengths we choose the optimal alignment according to the Bragg condition related to the average wavelength $\overline{\lambda} = (0.65 + 0.55)/2 = 0.6 \,\mu\text{m}$ which gives

$$\theta_{\rm B} = \frac{0.6 \times 10^{-6} \times 80 \times 10^6}{2 \times 620} = 0.0387.$$

The first diffracted beam corresponding to λ_1 is directed at the angle

$$\alpha_1 = \theta_{\rm B} + \frac{\lambda_1 f}{V_{\rm s}} = 0.0387 + \frac{0.65 \times 10^{-6} \times 80 \times 10^6}{620} = 0.12257$$

and the same consideration for the second wavelength yields

$$\alpha_2 = \theta_{\rm B} + \frac{\lambda_2 f}{V_{\rm s}} = 0.0387 + \frac{0.55 \times 10^{-6} \times 80 \times 10^6}{620} = 0.1097.$$

These two angles define the distance between the CCD centers in the image plane M behind lens L₂. Hence, assuming the light is parallel between two lenses and the plane M is the focal plane of L₂, we have for the lens focal length $\Delta \alpha \times f'_2 = (2.4 + 2) = 4.4$; $f'_2 = 4.4/(0.12257 - 0.1097) = 341$ mm. Therefore, the locations of both CCDs in M are as follows: $H_1 = \alpha_1 \times f'_2 = 0.12257 \times 341 = 41.80$ mm; $H_2 = \alpha_2 \times f'_2 = 0.1097 \times 341 = 37.41$ mm.

(b) We assume that the AOM is operated in the linear range and therefore the ratio of the efficiency for both wavelengths is $\eta_1/\eta_2 = (\lambda_2^2/\lambda_1^2) = (0.55/0.65)^2 = 0.716 = I_{1\lambda_1}/I_{1\lambda_2}$. The output CCD signals can be found as $i_{d1}/i_{d2} = (I_{1\lambda_1}R_1/I_{1\lambda_2}R_2)$, where R_1 and R_2 are the responsivity of the CCDs at both wavelengths. To calculate them we use Eq. (4.1) which gives $R_1/R_2 = (0.65 \times 0.32)/(0.55 \times 0.21) = 1.8$ and therefore $i_{d1}/i_{d2} = 0.716 \times 1.8 = 1.29$.

7.9. (a) Referring to Fig. 7.17, we find first the angle between the two receivers: $\Delta \alpha = l/L = 100/2,000 = 0.05$, and the frequency variation required for scanning at this angle: $\Delta \alpha = (\lambda/V_S) \times \Delta f$; $\Delta f = (0.05 \times 620/1.064 \times 10^{-6}) = 29.13$ MHz. Hence, communication is operated at RF signals of 50 ± 14.57 MHz.

(b) Due to diffraction of the laser beam inside the AOM the minimum angular width of a single Gaussian beam is

$$\delta\theta = 1.22 \frac{\lambda}{D} = 1.22 \frac{1.064 \times 10^{-3}}{1} = 1.298 \text{ mrad.}$$

Hence, to avoid cross-talks between two close receivers the angular distance between them should be equal to $\delta\theta$ and therefore the number of independent receivers is $N = \Delta \alpha / \delta\theta = 0.05/(1.298 \times 10^{-3}) = 38$.



FIGURE 7.17 Problem 7.9 – Schematic of communication system with two receivers.

7.10. From the efficiency graph shown in Fig. 7.10 we choose the interval where the efficiency is at least 50% of the maximum value. This gives $\Delta f = (120-40) = 80$ MHz and therefore the full angular range is

$$\Delta \theta = \frac{\lambda}{V_{\rm S}} \times \Delta f = \frac{0.6 \times 10^{-6}}{4,200} 80 \times 10^6 = 0.0114.$$

The diffraction limit yields $\delta\theta = (1.22 \times 0.6 \times 10^{-3}/5) = 0.1464 \times 10^{-3}$. Finally we get

$$N = \frac{\Delta\theta}{\delta\theta} = \frac{0.0114}{0.1464 \times 10^{-3}} = 78.$$

7.11. (a) We are interested in the maximum light beam size in the direction of the traveling acoustic wave, and therefore the laser diode optics should be positioned in such a way that the maximum diameter of the elliptical shape is parallel to the Bragg cell crystal. Then the 10 mm beam size dictates the time of interaction between the acoustic wave and the light, and the TBW is determined as follows (see Eq. (7.7)):

TBW =
$$\frac{D}{V_{\rm S}}\Delta f = \frac{10 \times 10^{-3}}{620} \times 40 \times 10^6 = 645.$$

Therefore, the maximum resolvable number of separated light spots is N = 645 and the minimum resolvable spectral interval is $\Delta f/N = (40 \times 10^6/645) = 62$ kHz, which is the spectral resolution of the system.

(b) Changing the light source does not affect the TBW if the shape of the beam at the Bragg cell entrance remains the same as with the original laser. Hence, the spectral resolution of the system after replacement of the laser will also remain the same, 62 kHz.

7.12. In the system for spectral measurement the location of the detector can be adjusted in such a way that each element of the array is responsible for a separate spectral interval, so that the total spectral range of the tested signals is $\Delta f = 1,024 \times 30 \text{ kHz} = 30.7 \text{ MHz}$. The time of interaction of light with the acoustic wave can be found from TBW = N = 1,024: $\tau = N/\Delta f = 1,024/(30.7 \times 10^6) = 33.35 \mu s$. This value dictates the size of the light beam inside the Bragg cell, if the acoustic velocity is known. Trying first PbMoO₄ material with $V_S = 4,200 \text{ m/s}$ we obtain $D = V_S \times \tau = 4,200 \times 33.35 \times 10^{-6} = 140 \text{ mm}$, which is not realistic. For the same reason a great number of acousto-optical materials with relatively high acoustic velocity cannot be exploited in the system. From those that have a low V_S we try TeO₂ in shear operation mode: $V_s = 620 \text{ m/s}$. The corresponding size D becomes $D = 620 \times 33.35 \times 10^{-6} = 20.7 \text{ mm}$, which is quite possible.

The next step is to choose a light source and lens. The diffraction limit of the system should result in a spot as small as a single pixel of the CCD, at least. Denoting the focal length of the lens as F, we have

$$\delta\theta \times F = \frac{\lambda}{D}F = 0.015 \text{ mm}; \quad \lambda F = 0.015 \times 20.7 = 0.3105 \text{ mm}^2.$$

Trying a laser of 0.6 μ m wavelength (He–Ne laser or a laser diode), we get $F = 0.3105/(0.6 \times 10^{-3}) = 517.5$ mm. If a Ga–As laser diode of 0.83 μ m wavelength is chosen then $F = 0.3105/(0.83 \times 10^{-3}) = 374$ mm. We prefer the latter case since it gives a shorter optical path.

The last step is to choose the anamorphic collimation optics. As discussed in Chapter 3, an anamorphic ratio of 1:3 to 1:6 can be easily achieved by using an anamorphic prism pair. Depending on the ellipticity of the beam at the laser diode output one can vary the prism pair in order to get the reasonable ratio of 1:5 after the collimation lens, just at the entrance to the AOM. Hence, the Bragg cell required for the system should be a TeO₂ crystal of 21 mm in length by 4 mm in height. It is followed by a lens of 374 mm focal length and is illuminated by a Ga–As laser diode with anamorphic collimation optics providing a 1:5 elliptical beam.

Optical Systems for Distance and Size Measurements

8.1. Laser Rangefinders

Rangefinders are instruments intended for distance measurements, usually (but not necessarily) in the open air. A schematic of a simple configuration is depicted in Fig. 8.1. The concept of measurement is quite simple. The light pulses of a laser source travel over a range *S* to a target and then, after reflection on the target surface, come back to the transmitter optics where they are received and analyzed by the detector circuitry in order to measure the total time of flight, t_f . The range *S* is then calculated as

$$S = \frac{vt_{\rm f}}{2} = \frac{c}{n} \frac{t_{\rm f}}{2} \tag{8.1}$$

where v is the velocity of light propagation in the medium (in the air), $c = 3 \times 10^8$ m/s is the velocity of electromagnetic waves in a vacuum, and n is refractive index of air. Lenses L₁ and L₂ provide a means of expanding the laser beam and reducing the beam divergence (see Section 3.3.2). As is evident from Eq. (8.1), the accuracy of measurement as well as the limitations on maximum and minimum measured range depend on: (i) the processing ability of the detector circuitry; (ii) the variation of refractive index (due to temperature change, gradients of density, wind, etc.); and (iii) the variation of the optical path (due to random fluctuations in the atmosphere).

When an emerging laser pulse leaves the transmitter a synchronizing pulse is registered by the processing electronics. Then, the time interval t_f is measured by comparing with the circuitry clock the rise front of the synchronizing pulse



FIGURE 8.1 (a) Schematic of a laser rangefinder and (b) a sequence of light pulses.

and the rise front of the detector signal originating in the light pulse returned from the target. The shorter the laser pulses and the greater the electronic bandwidth the smaller the uncertainty in the time of flight measurement. Another important issue is related to the reflective properties of the target. Usually it is assumed that the target surface is an ideal diffuser scattering the reflected light uniformly in a hemisphere. Details related to the detector signals are considered in Problems 8.1 and 8.2.

The refractive index of air as a function of density obeys the Gladston–Dale formula (Eq. (7.1); see Section 7.2.1) $(n - 1)/\rho = K$ with the constant $K = 0.226 \text{ cm}^3/\text{g}$. Since the density, ρ , of air varies with temperature as $\rho = \rho_0(1 + \beta t_C^0)$, where $\rho_0 = 0.001293 \text{ g/cm}^3$ is the density under normal conditions ($t_C^0 = 0_C^0$; P = 760 mmHg), we have for the refractive index

$$n = 1 + K\rho_0(1 + \beta t). \tag{8.2}$$

Furthermore, assuming that air obeys the ideal gas relations which yields $\beta = 1/T_{\rm K}^0$, one obtains the variation of refractive index with temperature in the following form:

$$dn = K \rho_0 \beta dT = 0.000292 \frac{1}{T} dT.$$
 (8.3)

As to the optical path variation caused by the random fluctuations of the atmosphere (turbulence), we mention here that the light pulses propagating through the turbulent atmosphere do not strictly travel along a straight line connecting the transmitter and the target, but travel according to randomly variable trajectory the total length of which depends on statistical parameters of the turbulence. This is quite a complicated phenomenon the description of which is beyond the scope of this book.

Problems

8.1. Find the minimum distance which can be measured by a rangefinder operated with laser pulses of 20 ns duration and detector electronics of 100 MHz bandwidth.

8.2. A laser rangefinder comprises a Nd:YAG laser (wavelength 1.06 μ m) which generates light pulses of 0.5 mJ energy and 20 ns duration at a repetition rate of 10 pulses/s, a silicon detector (quantum efficiency 0.58) with electronic circuitry of 10 MHz bandwidth and 1 nA dark current, and transmitting optics of 3 cm diameter. Assuming the target reflectance is 0.4, find the maximum range measured by the device.

8.3. At 6:00 in the morning a distance *S* is measured using a laser rangefinder and the result is 8,100 m. At 13:00 when a temperature rise of 15° is experienced a repeat measurement of *S* is carried out. Assuming the laser generates light pulses of 30 ns and the uncertainty in the measurement of the flight time is 2% of the pulse duration, check if the measurement results are affected by the temperature change.

8.2. Size Measurement with a Laser Scanner

Of numerous possible architectures we consider a configuration with a fast-rotating mirror (like the scanner depicted in Fig. 7.2). The arrangement of a system intended for the measurement of linear dimensions is illustrated in Fig. 8.2. The object to be tested is a rod of diameter D. A mirror M is rotated around the horizontal axis at constant rotation speed, ω . A laser beam is aligned along the axis of mirror rotation in such a manner that point A where the beam strikes the mirror remains



FIGURE 8.2 (a) Schematic of size measurement with a fast-rotating scanner and (b) the time history of the detector signal.

unchanged and does not move along the mirror surface. This point A is the front focus of lens L_1 and, as a consequence, the scanning beam reflected by the rotating mirror moves in the vertical plane and behind the lens it remains parallel to the lens axis. Lens L_2 collects the incident light and transfers it to a detector S located at the back focus of the second lens.

The scanning range of the beam between the lenses is larger than the measured size (diameter D). Thus, the detector signal varies during beam scanning from its maximum value, i_{max} , to zero, when it is obstructed by the test body, and then it returns to its initial (maximum) value. The processing electronics measure the time interval, τ , when the beam is obstructed (see Fig. 8.2b) giving the diameter size as $D = V\tau$, where V is the linear speed of the beam motion between the lenses. Since $V = f'_1 \omega$, one can finally get the working formula

$$D = f_1' \omega \tau. \tag{8.4}$$

The accuracy of the measurement depends on the detector properties, on the speed errors, and on the laser beam shape and stability. Details of accuracy considerations are given in the solution to Problem 8.4.

Problems

8.4. An optical system for rod size measurement includes a fast-rotating scanner (rotation speed $\omega = 6,000$ rpm), a He–Ne laser of 2 mW power, 2 mrad divergence angle and 0.6 mm beam size at the cavity exit, two identical lenses of 100 mm focal length and f# = 2.0, and a silicon p–i–n photodiode with NEP = 7×10^{-6} W/Hz^{-1/2} and $\eta = 0.8$. The laser stability is 1% and signal processing is performed by 8-bit digital electronic circuitry (255 discrimination levels).

- (a) Find the linear dynamic range required for the detector.
- (b) Assuming that the instability of the rotation speed is 1% and the uncertainty of the focal lengths is 2% calculate the overall accuracy of the rod diameter measurement.
- (c) Show how the accuracy can be improved by calibration of the device using a rod of well-known diameter $D_1 = (5 \pm 0.002)$ mm.

8.3. Interferometric Configuration

A laser interferometer is one of the most accurate tools for linear displacement measurement. Although a great number of possible configurations have been



FIGURE 8.3 Displacement measurements with a laser interferometer.

developed and successfully explored, the same basic approach is always utilized. That is, the displacement to be measured causes an optical path difference between the reference branch and the working branch of the interferometer which results in an oscillation of interference intensity registered by the detector circuitry. Each period of oscillation corresponds to a very small displacement of a half wavelength, so that simple counting of the number of sequential oscillations can be easily and precisely interpreted as an overall accumulated displacement. A schematic of a typical laser interferometer is illustrated in Fig. 8.3.

The key part of the system is an interferometer, T, which consists of a cubic beam splitter with a 90° prism. A laser beam is split inside this interferometer into two parts: one goes to the prism and then back to the beam splitter and further to a detector D (the reference branch); the other goes to a retroreflector, R, connected to a test object moving along a surface A (displacement branch) and then goes back to the beam splitter and proceeds to the detector where it undergoes interference with the reference beam. The measured displacement *S* is traveled twice by the laser beam and therefore the total number of oscillation, *N*, of the detector signal is related to the displacement, *S*, and the laser wavelength, λ , as follows:

$$S = \frac{\lambda}{2}N.$$
(8.5)

The measurement errors can be of different origins. We address here the errors caused by misalignment of the laser beam to surface A. Let a small tilt, θ , exist between the traveling path and the laser optical axis (Fig. 8.4). Then a retroreflector, R, connected rigidly to a moving body, B, participates simultaneously in three motions: (i) translation along the horizontal axis (apparent measured displacement, *S*); (ii) translation, δ , in the vertical direction; and (iii) rotation around the horizontal axis perpendicular to the cross-section of the system (not shown on the figure). The third motion is not important, since the incident beam and the exit beam of the retroreflector are always parallel to each other (as long as the retroreflector vortex angle does not differ from 90°). The first two motions cause the so-called



FIGURE 8.4 Misalignment of a laser beam to a motion surface A: (a) origin of the cosine error; (b) origin of the signal contrast reduction.

"cosine error," which is the difference between the calculated distance *S* and the real translation S':

$$\Delta S = S' - S = S(1 - \cos \theta). \tag{8.6}$$

Besides this, translation δ in the vertical direction reduces the overlapping fraction of the interfering beams by an amount 2δ (see Fig. 8.4b). As a result, the contrast of the detector signal oscillations (defined as $C = (i_{det}^{(max)} - i_{det}^{(min)})/(i_{det}^{(max)} + i_{det}^{(min)}))$ is also reduced. Consideration of the above errors in more detail can be found in the solutions to Problems 8.5 and 8.6.

Problems

8.5. An interferometric system for distance measurements like that of Fig. 8.3 is operated with a He–Ne laser beam expanded to 30 mm in size and transferred through a stop of 10 mm in diameter located at the entrance of the interferometer (in order to minimize the influence of intensity reduction in the radial direction).

- (a) Calculate the number of counts registered by the system electronics when the measured displacement is S = 1,200 mm.
- (b) A 75% reduction of the signal contrast is experienced when the retroreflector is displaced to the far end of the measured range. Find the misalignment of the laser to the motion surface and calculate the cosine error in this case.

8.6. The light source exploited in the system shown in Fig. 8.3 is a He–Ne laser of 400 mm cavity length. Assuming the aperture stop near the detector is large enough (does not truncate the laser beams) and the minimum acceptable contrast



FIGURE 8.5 (a) Schematic of 3-D shape measurements with stratified light and (b) the grid image in the CCD plane.

of the detector signal is $C_{\min} = 10\%$, what is the maximum displacement that can be measured with the system?

8.4. Stratified Light Beam and Imaging Measuring Technique

Measurement of the 3-D shape of a body is a frequently encountered problem in numerous application areas. In many cases, especially in industrial environments where productivity is a crucial issue, methods based on imaging can be good solutions. Two approaches can be effectively realized. The first exploits point-by-point scanning of a measured surface by a laser beam. At each location the point of intersection of the beam with the studied surface is imaged on a 2-D detector (usually a CCD) and analyzed by an image processor. Actually what is involved here is a simple triangulation (see details in Problem 8.7).

The other approach is based on so-called stratified light illumination when the illumination beam creates on the measured surface a line of light or a 1-D or 2-D grid of light lines. The main idea is demonstrated in Fig. 8.5. A surface P described as z = F(x, y) is illuminated by radiation emitted from a light grid generator (1) and a set of vertical and horizontal lines is projected on the surface. An imaging module (2) creates the image of the surface with the light grid in the plane of a CCD. Deformation of the grid image lines (clearly seen in Fig. 8.5b) reflects the influence of the variation of the heights *z* along the *x* and *y* coordinates of the surface P. The interpretation of the image structure (again, based on the triangulation principle) allows for the restoration procedure and image processing



FIGURE 8.6 Problem 8.7 – Geometry of rays in a system for vertical distance measurement (optical profiler).

with sub-pixel accuracy (see explanation in Chapter 4, Problems 4.16 and 4.17) is required very frequently.

Problems

8.7. A distance measurement system based on imaging comprises a He–Ne laser, a lens L of 10 mm focal length, and a CCD detector with 640 \times 512 pixels, 7 \times 7 μ m each. The basis segment B (see Fig. 8.6) is of 30 cm and the observation angle $\alpha = 18^{\circ}$.

- (a) Calculate the maximum and minimum distances which can be measured by the system.
- (b) Assuming that image processing allows one to reveal minimum variations of a pixel size, find the accuracy of the vertical distance measurement.

8.8. Gear teeth testing is performed with a light grid of 3 by 3 lines, 0.5 mm width each, and 1.5 mm spacing. The imaging branch includes a lens L of 40 mm focal length and a CCD detector of 4.8 mm \times 5.6 mm area, 8.3 μ m pixel size. The maximum size of the gear tooth is 5 mm by 10 mm (see Fig. 8.7) and it is measured in a single shot (one frame of the CCD). Is it possible to achieve an accuracy of measurement of as high as 0.01 mm? How should the system be arranged in such a case?



FIGURE 8.7 Problem 8.8 - (a) Configuration of a system for gear profile measurement and (b) grid image in the CCD plane.

8.5. Solutions to Problems

8.1. The minimum time of flight t_f is achieved if the outgoing pulse is followed immediately by the back reflected pulse, i.e., $t_f = \tau = 20$ ns. The bandwidth of 100 MHz means that the minimum time interval between two events which can be processed separately by the system electronics is $\Delta t = 2/(100 \times 10^6) = 20$ ns, i.e., compatible with t_f . Hence, from Eq. (8.1) we get

$$S_{\min} = \frac{ct_{\rm f}}{2} = \frac{3 \times 10^8 \times 20 \times 10^{-9}}{2} = 3 \,\mathrm{m}.$$

8.2. We find first the number of photons in the laser pulse. As a single photon of $1.06 \,\mu\text{m}$ wavelength has an energy of

$$\frac{hc}{\lambda} = \frac{6.63 \times 10^{-34} \times 3 \times 10^8}{1.06 \times 10^{-6}} = 1.88 \times 10^{-19} \,\mathrm{J}$$

the total number of photons in the pulse is $N_p = 5 \times 10^{-4}/1.88 \times 10^{-19} = 2.66 \times 10^{15}$ photons. Of that amount a fraction $RN_p = 0.4N_p$ is reflected by the target in all directions (in a solid angle of 2π) of which the fraction

$$\frac{\pi D^2}{4S^2} \frac{1}{2\pi} = \frac{D^2}{8S^2}$$

is reflected in the direction of the rangefinder and captured by its optics. If we also take into account the transmittance of the atmosphere while the light travels twice the distance *S*, $T = \exp(-2\alpha S)$, we find the total number of photons, N_d , striking the system detector as follows:

$$N_{\rm d} = 0.4N_{\rm p} \frac{D^2}{8S^2} e^{-2\alpha S} = 0.4 \times 2.66 \times 10^{15} \frac{9 \times 10^{-4}}{8S^2} e^{-2\alpha S} = \frac{1.19 \times 10^{11}}{S^2} e^{-2\alpha S}.$$

Keeping in mind the quantum efficiency of the detector, $\eta = 0.58$, we find also the number of electrons generated in the detector by radiation reflected from the target:

$$N_{\rm e} = 0.58 N_{\rm d} = \frac{0.69 \times 10^{11}}{S^2} e^{-2\alpha S}.$$

As the minimum time interval registered by the electronic circuitry is $\tau = 1/\Delta f = 10^{-7}$ s, one can find the detector signal (current), i_{det} , originating in the captured pulse:

$$i_{\text{det}} = \frac{0.69 \times 10^{11}}{S^2} e^{-2\alpha S} \frac{1.6 \times 10^{-19} c}{10^{-7} \text{ s}} = 0.11 \frac{e^{-2\alpha S}}{S^2} \text{ A}$$

Furthermore, we assume that the minimum signal-to-noise ratio (SNR, see Section 4.1) should be 3:1 at least and that the noise of the detector is due to dark current primarily, $i_{d.c.} = 1$ nA. Then we get the following equation for the maximum range S: SNR = $3 = i_{det}/i_{d.c.} = 1.1 \times 10^8 (e^{-2\alpha S}/S^2)$ or, remembering that $\alpha = 0.1$ km⁻¹ = 10^{-4} m⁻¹ and taking the square root and the logarithm of both sides:

$$S = e^{-\alpha S} 0.609 \times 10^4; \quad \ln S = 8.714 - 10^{-4} S.$$
 (A)

Since the last term is small, the transcendental equation (A) can be easily solved by iteration (or by a trial and error method) which finally yields $S_{\text{max}} = 4,100$ m.

8.3. From Eq. (8.3) we get the change of refractive index due to a temperature rise of 15° : $dn = 0.000292 \times (1/300) \times 15 = 1.46 \times 10^{-5}$. Then, taking the logarithm derivative of both sides of Eq. (8.1) and assuming that $t_{\rm f}$ remains the same in the morning and in the afternoon, we have

$$\Delta S = -S \frac{\Delta n}{n} = 5,000 \frac{1.46 \times 10^{-5}}{1.000292} = 0.1168 \text{ m}.$$

This change of the calculated range S is equivalent to a change of the flight time of

$$\Delta t_{\rm f} = \frac{2 \times \Delta S}{c} = \frac{2 \times 0.1168}{3 \times 10^8} = 0.78 \text{ ns}$$

which is greater than the uncertainty of the laser pulse duration (0.6 ns) and therefore can be resolved by the system processor. Thus, the measurement result is affected by the temperature variation, meaning that a correction from the environment temperature is required. **8.4.** (a) We address the system configuration as presented in Fig. 8.2 and assume that the waist of the laser beam is 0.6 mm and it is located on the rotating mirror surface. Then after lens L_1 the laser beam waist remains of the same size, *w*, and it is located at a distance S' = f' behind the lens (see explanation in Section 3.3.2, Eq. (3.7)). The same consideration is also valid for the second lens and therefore the two lenses should be positioned 200 mm from one another and detector S is to be located 100 mm behind lens L_2 .

The dynamic range of the detector is defined as follows (see Eq. (4.5)):

$$DR = \frac{i_{det.max}}{i_{det.min}} = \frac{P_{max}}{P_{min}}$$
(A)

where P_{max} is the maximum radiation power incident on the detector when it is not saturated and the light beam is not shaded by the studied body and P_{min} is the minimum detected radiation power (this is achieved when the light beam is almost totally shaded by the measured body). It is apparent that $P_{\text{max}} = 2$ mW. As to the second value, P_{min} , one can calculate it in terms of noise equivalent power (NEP) and the system bandwidth (Δf): $P_{\text{min}} = \text{NER}\sqrt{\Delta f}$. To find Δf we consider the dynamics of the laser beam motion relative to the measured rod (see Fig. 8.8a), keeping in mind that the full size of the beam is about three times larger than the beam waist, $3 \times 0.6 = 1.8$ mm, and the linear velocity of the beam moving between two lenses is $V = \omega F$ (where ω is the angular velocity of rotation and F = 100 mm is the focal length of each lens). Hence

$$\tau_{\rm tr} = \frac{1.8}{\omega F} = \frac{1.8 \text{ mm}}{100 \text{ s}^{-1} \times 100 \text{ mm}} = 180 \text{ }\mu\text{s}.$$

Thus, for 8-bit electronic circuitry the minimum resolvable time interval should be $\tau_{\rm m} = \tau_{\rm tr}/255 = 0.706 \ \mu s$ and the Nyquist theorem gives the necessary bandwidth as

$$\Delta f = \frac{1}{2\tau_{\rm m}} = \frac{10^6}{2 \times 0.706} = 0.708 \text{ MHz}.$$

Therefore, $P_{\min} = 7 \times 10^{-16} \sqrt{0.708 \times 10^6} = 5.89 \times 10^{-13}$ W. By substituting this value in Eq. (A) we finally get $DR = (2 \times 10^{-3})/(5.89 \times 10^{-13}) = 3.33 \times 10^9$.

(b) Given the angular velocity of the scanner and the focal length of the lenses, measurement of the rod diameter D is based on measurement of the transition time interval, τ , shown in Fig. 8.8a:

$$D = V\tau = \omega F\tau. \tag{B}$$

From Eq. (B) one can estimate the accuracy of measurement in a standard way:

$$\frac{\Delta D}{D} = \frac{\Delta \tau}{\tau} + \frac{\Delta \omega}{\omega} + \frac{\Delta F}{F}.$$
 (C)



FIGURE 8.8 (a) Motion of a laser beam and corresponding signal of a detector and (b) geometry of the beam at the measured rod.

The last term in this expression is constant for a given system and it can be avoided by proper calibration or measuring the lens focal distance. The second term on the right-hand side does not depend on electro-optical elements and can be treated by electronic and electro-mechanical means. The first term on the right-hand side is dictated by the random noise of the laser and of the detector and it is our main concern here. Referring to the upper graph of Fig. 8.8a, we realize that the uncertainty in the measurement of τ is affected by the transition function of the system, $\Phi(t) = i_{det}(t)$, and by its derivatives at the time moments t_1 , and t_2 , since $\delta \tau = \delta i/(d\Phi/dt)_{t=t_1}$ (we assume that the same relation is valid for the time moment t_2). To find the function $\Phi(t)$ we consider the radiation power, *E*, incident on the detector when the center of the laser beam is at a distance *y* from the rod wall and a fraction of the beam is shaded by the rod (for simplicity we consider the rod wall as a straight surface because its radius of curvature is much greater than that of the light beam). Considering Figure 8.8b, we get

$$E = P_{\max} - 2I_0 \int_0^{\alpha} \int_{y/\cos\varphi}^R \exp\left(-\frac{2r^2}{w^2}\right) r \, dr d\varphi$$

= $P_{\max} + I_0 \alpha \frac{w^2}{2} e^{-(2R^2/w^2)} - I_0 \frac{w^2}{2} \int_0^{\alpha} \exp\left(-\frac{2y^2}{w^2\cos^2\varphi}\right) d\varphi.$

In this expression we use the laser beam profile described by Eqs. (3.1) and (3.2) with the center light intensity I_0 as per Eq. (3.6). Taking into account that $\Phi = R_{\lambda}E$, where R_{λ} is the responsivity of the detector defined in Section 4.1 and equal in our case to $\eta \lambda / 1.24 = 0.8 \times 0.63 / 1.24 = 0.4$ A/W, we obtain

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \frac{\mathrm{d}\Phi}{\mathrm{d}y/V} = R_{\lambda}\omega F \frac{\mathrm{d}E}{\mathrm{d}y} = R_{\lambda}\omega F I_0 \frac{w^2}{2} \int_0^{\omega} \frac{\mathrm{d}}{\mathrm{d}y} \left[\exp\left(-\frac{2y^2}{w^2\cos^2\varphi}\right) \right] \mathrm{d}\varphi$$
$$= 2R_{\lambda}\omega F I_0 y U \tag{D}$$

where the following integral is introduced:

$$U = \int_{0}^{\alpha} \exp\left(-\frac{2y^2}{w^2\cos^2\varphi}\right) \frac{\mathrm{d}\varphi}{\cos^2\varphi} = \int_{0}^{\tan\alpha} \exp\left[-\frac{2y^2(1+z^2)}{w^2}\right] \mathrm{d}z$$
$$= \sqrt{\frac{\pi}{2}} \frac{w}{2y} e^{-\frac{2y^2}{w^2}} \operatorname{erf}\left(\sqrt{2}\frac{y}{w}\tan\alpha\right) \tag{E}$$

expressed in terms of the error function

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_{0}^{z} e^{-x^{2}} \mathrm{d}x.$$

By substituting U in Eq. (D) and keeping in mind that

$$\frac{y}{w}\tan\alpha = \frac{y}{w}\frac{\sqrt{R^2 - y^2}}{y} = \frac{\sqrt{R^2 - y^2}}{w}$$
 (F)

we obtain

$$\left(\frac{\mathrm{d}\Phi}{\mathrm{d}t}\right)_{t_1} = R_\lambda \omega F I_0 \sqrt{\frac{\pi}{2}} \operatorname{werf}\left(\sqrt{2}\frac{R}{w}\right) \tag{G}$$

where it is taken into account that at $t = t_1 y = 0$. Investigating the behavior of the error function term in Eq. (G) at different ratios R/w we find: for R/w = 1, erf $(\sqrt{2}) = 0.953$; for R/w = 2, erf $(2\sqrt{2}) = 0.99992$; for larger R/w it is equal to 1.000. Hence, Eq. (G) gives

$$\left(\frac{\mathrm{d}\Phi}{\mathrm{d}t}\right)_{t_1} = R_\lambda \omega F P_{\max} \sqrt{\frac{2}{\pi}} \frac{1}{w} = 0.798 \frac{R_\lambda \omega F P_{\max}}{w}$$

and therefore

$$\delta \tau = \frac{w}{0.798 R_{\lambda} \omega F P_{\text{max}}} \delta i = \frac{1.25 w}{R_{\lambda} \omega F P_{\text{max}}} \delta i \tag{H}$$

where Eq. (3.6) is used for I_0 .

The uncertainty in the detector current due to shot noise of the detector is governed by Eq. (4.11). We apply it for the current value $i_{det max}/2$ (see Fig. 8.8a) where the maximum current of the detector is defined as $i_{det max} = R_{\lambda}P_{max} = 0.4A/W \times 2mW = 800\mu$ A.Thus, $\delta i_{Sn} = \sqrt{2 \times 1.6 \times 10^{-19} \times 400 \times 10^{-6} \times 0.708 \times 10^{6}} = 9.5 \times 10^{-9}$ A. The uncertainty of the detector current due to the laser intensity noise is calculated as follows: $\delta i_{L} = R_{\lambda} \times \Delta P = 0.4 \times 0.01 \times 2 \times 10^{-3} = 8 \times 10^{-6}$ A. Thus δi_{L} definitely dominates over the shot noise. Hence, from Eq. (H) we get

$$\delta \tau = \frac{1.25 \times 0.3 \times 8 \times 10^{-6}}{800 \times 10^{-6} \times 10^4} = 0.375 \,\mu \text{s}.$$

The total uncertainty of time measurement is $\Delta \tau = 2 \times \delta \tau = 0.75 \,\mu s$. By substituting this value in Eq. (C) we find

$$\Delta D = \Delta \tau \times \omega F + D\left(\frac{\Delta \omega}{\omega} + \frac{\Delta F}{F}\right) = 0.75 \times 10^{-6} \times 10^4 + D(0.01 + 0.02)$$
$$= 7.5 \,\mu\text{m} + 0.03D$$

which renders for the maximum measured diameter (D = 48.8 mm) $\Delta D = 1.47 \text{ mm}$; $\Delta D/D = 3.0\%$.

(c) A rod of well-known diameter D_1 is used for calibration of the system. Since the product ωF remains the same in both situations, in calibration and in normal operation, one can improve measurement accuracy by performing a relative measurement procedure: $D/D_1 = \tau/\tau_1$; $\Delta D/D = \Delta D_1/D_1 + \Delta \tau/\tau + \Delta \tau_1/\tau_1$. For $D_1 = 5$ mm we get $\tau_1 = 5 \times 10^{-4}$ s; and $\Delta \tau_1/\tau_1 = 0.75 \times 10^{-6}/5 \times 10^{-4} = 0.15\%$ and therefore $\Delta D = 0.002(D/D_1) + 0.0075 + 0.0015D$ which yields for the maximum diameter D = 48.8 mm $\Delta D = 0.1$ mm; $\Delta D/D = 0.2\%$.

8.5. (a) Using Eq. (8.4) one can find the number of counts registered by the detector while the test body is displaced by S = 1,200 mm: $N = 2S/\lambda = 2,400/0.63 \times 10^{-3} = 3,809,523$ counts.

(b) We refer to Fig. 8.4 and assume that the total power of the laser, P, is divided equally between two interfering channels, P/2 each. Due to misalignment error only a fraction of the power in each channel participates in interference (we denote this fraction as $q(\delta)$ and it is related to the shaded area indicated in Fig. 8.4b). Hence, the oscillating part of the power incident on the detector is changed from the maximum value of $P_{\text{max}} = 2Pq$ to the minimum (zero) value of $P_{\text{min}} = 0$. The rest of the power from both channels, P(1 - q), comes to the detector with no oscillation. Therefore, the maximum total power coming to the detector is $I_{\text{tot.max}} = 2Pq + P(1 - q) = P(1 + q)$ and the minimum total power is $I_{\text{tot.min}} = P(1 - q)$, and the contrast of the detector signal can be expressed

as follows:

$$C = \frac{I_{\text{tot.max}} - I_{\text{tot.min}}}{I_{\text{tot.max}} + I_{\text{tot.min}}} = \frac{P(1+q) - P(1-q)}{P(1+q) + P(1-q)} = q.$$
 (A)

Since the intensity variation inside the beam is negligible, one can assume that the total power incident on an area is a linear function of the area size. The shaded area shown in Fig. 8.4b depends on translation δ caused by misalignment error, or, more specifically, on the angle α (shown in the figure) and the ratio δ/R , where *R* is the radius of the system stop. A simple geometrical consideration yields

$$A = \alpha R^2 - R^2 \sin \alpha \times \cos \alpha = \alpha R^2 - \delta R \sqrt{1 - (\delta/R)^2}$$
(B)

where $\cos \alpha = \delta/(R)$ and therefore

$$q = \frac{A}{\pi R^2} = \frac{1}{\pi} \left[\alpha - \frac{\delta}{R} \sqrt{1 - (\delta/R)^2} \right] = C.$$
 (C)

Given the contrast value, Eq. (C) is a non-linear equation with regard to δ . In our case we have C = q = 0.25 and by trial and error (or using a simple iteration process) we obtain $\delta/R = 0.4$ which gives $\delta = 0.4 \times 5 = 2.0$ mm. Thus, the misalignment angle, θ , is calculated as $\theta = \delta/S = 2/1,200 = 1.7 \times 10^{-3} = 0.09^{\circ}$ and the cosine error is $\Delta S = 1,200(1 - \cos 0.09^{\circ}) = 0.0016$ mm.

8.6. We start with the calculation of the laser beam parameters using expressions of Section 3.3.1. From Eq. (3.4) we find the beam waist inside the cavity:

$$w_0 = \sqrt{\frac{\lambda L}{2\pi}} = \sqrt{\frac{0.63 \times 10^{-3} \times 400}{2\pi}} = 0.2 \text{ mm}$$

which enables one to calculate the divergence angle of the beam as per Eq. (3.3): $2\theta = 2(0.63 \times 10^{-3})/(\pi \times 0.2) = 2 \times 10^{-3}$. The beam size at the entrance of the laser is $2w_1 = 2w_0\sqrt{2} = 0.566$ mm and it is supposed to be very close to the interferometer. Thus, the first channel provides to the detector practically a beam of size $2w_1$. The second channel projects on the detector a significantly larger beam, since it is divergent, after traveling twice the distance *L*, of size $2w_2 = 2\sqrt{w_0^2 + 4\theta^2 L^2}$ (see Eq. (3.1)). The intensity of both interfering beams is also different. The values at the center of the beams, I_{01} and I_{02} , obey Eq. (3.6) and therefore are related as follows: $I_{01}/I_{02} = (w_2/w_1)^2$. The above discussion is summarized in Fig. 8.9 where two beams are shown at the entrance to the detector aperture stop. Keeping in mind that the oscillating signal of the detector is caused by the fraction of the area where the two beams interfere with one another, i.e., inside the circle



FIGURE 8.9 Problem 8.6 – Two beams as they arrive at a system detector. D' (dotted line) is the aperture stop.

indicated as I_1 , we can find the maximum and minimum of the oscillating power as follows:

$$E_{\max} = \pi w_1^2 \left(\sqrt{I_1} + \sqrt{I_2} \right)^2 = \pi w_1^2 I_2 (k+1)^2;$$

$$E_{\min} = \pi w_1^2 \left(\sqrt{I_1} - \sqrt{I_2} \right)^2 = \pi w_1^2 I_2 (k-1)^2$$

where the ratio $k = w_2/w_1 = \sqrt{I_1/I_2}$ is introduced. Outside the area of I_1 there is no interference, but only a fraction of the second-channel beam (which we assume be equally spread over the area) is present, so that the optical power here is $E' = \pi (w_2^2 - w_1^2)I_2$. Hence, the contrast of the detector signal can be expressed as follows:

$$C = \frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min} + 2E'} = \frac{\pi w_1^2 I_2 [(1+k)^2 - (1-k)^2]}{\pi w_1^2 I_2 [(1+k)^2 + (1-k)^2] + 2\pi I_2 w_1^2 (k^2 - 1)}$$
$$= \frac{2k}{2k^2} = \frac{1}{k}.$$
 (A)

In our case C = 0.1; and $k = 10 = w_2/w_1$ and therefore

$$w_2 = 10w_1;$$
 $w_0^2 + 4\theta^2 L^2 = 100w_1^2 = 200w_0^2;$
 $L = \frac{w_0\sqrt{199}}{2\theta} = \frac{0.2\sqrt{199}}{2 \times 10^{-3}} = 1,411 \text{ mm}.$

8.7. (a) Keeping in mind that $|S| \gg S'$ we can replace in the calculation S' by f' and find the half-field angle, β , as follows (*N* is the total number of pixels in a line



FIGURE 8.10 Problem 8.7 – Consideration of measurement error in the vertical direction.

of the CCD):

$$\tan \beta = \frac{N}{2} \frac{\delta}{f'} = \frac{640 \times 7 \times 10^{-3}}{2 \times 10} = 0.224; \quad \beta = 12.62^{\circ}.$$

Furthermore, from the triangle O₁OQ (see Fig. 8.6) we get $|S| = B/\sin \alpha$ and from the triangle O₁OO₂ O₁O₂ = $|S| \sin \beta / \sin(\alpha + \beta)$ which gives

$$L_{\min} = O_1 Q - O_1 O_2 = \frac{B}{\sin \alpha} \left[\cos \alpha - \frac{\sin \beta}{\sin(\alpha + \beta)} \right]$$
$$= \frac{30}{\sin 18^{\circ}} \left(\cos 18^{\circ} - \frac{\sin 12.62^{\circ}}{\sin 30.62^{\circ}} \right) = 50.7 \text{ cm}$$
$$L_{\max} = O_1 Q + O_1 O_3 = \frac{B}{\sin \alpha} \left[\cos \alpha + \frac{\sin \beta}{\sin(\alpha - \beta)} \right]$$
$$= \frac{30}{\sin 18^{\circ}} \left(\cos 18^{\circ} + \frac{\sin 12.62^{\circ}}{\sin 5.38^{\circ}} \right) = 319.0 \text{ cm}.$$

(b) To find the error of measurement consider Fig. 8.10 where the segment M_2M_1 has a size of a single pixel, δ . Therefore, the angle φ which defines the error Δz in the vertical direction is calculated as $\tan \varphi = \delta/f' = 7 \times 10^{-4}$ and from the triangle $T_1T_2O_1$ (where the angle $O_1T_2T_1 = \alpha + \varphi$ and the second angle $T_2O_1T_1 = 90^\circ - \alpha$): $T_1O_1 = \delta/V = \delta B/(f' \sin \alpha)$ and

$$\Delta z = \frac{\delta}{V} \frac{\sin(90^{\circ} - \varphi)}{\sin(\alpha + \varphi)} = \frac{7 \times 10^{-3} \times 300}{10 \times \sin 18^{\circ}} \frac{\cos \varphi}{\sin 18.04^{\circ}} = 2.19 \text{ mm}.$$

8.8. Measurement of a whole tooth in a single shot requires that the image of a tooth be no larger than the CCD size. This dictates the magnification of the imaging optics and positioning of the lens (the distance *S*):

$$V = -\frac{5.6}{10} = -0.56;$$
 $S = f'\frac{1-V}{V} = 40\frac{1+0.56}{-0.56} = -11.4 \text{ mm}.$

Illuminator IS is tilted in the ZOY plane by an angle ψ and this angle does not affect the accuracy (it influences the width and spacing of the grid lines which are assumed to be on the measured surface as specified in the problem, i.e., 0.5 mm width and 1.5 mm spacing). The angle α in the ZOX plane determines the position of the imaging optics and it does influence the accuracy of measurements, as shown in Problem 8.7: $\Delta z = \delta/(V \sin \alpha)$. Obviously the greater the angle α the less the error Δz that can be achieved for a given pixel size ($\delta = 8.3 \times 10^{-3}$ mm) and magnification V. However, one should keep in mind that all lines of the grid have to be imaged with no overlapping in the CCD plane. In other words, two adjacent lines should be incident on two different lines of the CCD which means that $l_{\rm min} = 8.3 \times 10^{-3} \,\mathrm{mm} = (1.5 - 2 \times 0.25) V \cos \alpha$. This requirement dictates the maximum acceptable value of angle α : cos $\alpha_{max} = (8.3 \times 10^{-3})/0.56 = 0.01482$; $\alpha_{\rm max} = 89^{\circ}$ and therefore $\Delta z = (8.3 \times 10^{-3})/(0.56 \sin 89^{\circ}) = 14.8 \,\mu{\rm m}$, which does not meet the requirements of the problem. To achieve better results an image processing procedure with sub-pixel accuracy should be applied (see description in Section 4.4, Problem 4.17). In such a case the image of a single line of the grid (of 0.55 mm width) has to be projected on three sequential pixels at least. Then, we get $0.5V \cos \alpha' = 3 \times 8.3 \times 10^{-3}$; $\cos \alpha' = 0.0249/0.28 =$ 0.0889; $\alpha' = 84.9^{\circ}$ and a sub-pixel accuracy of 0.5 pixel (which can be easily achieved) is enough in order to get the measurement error smaller than 0.01 mm: $\Delta z = (0.5 \times 8.3 \times 10^{-3})/(0.56 \times \sin 84.9^{\circ}) = 7.4 \,\mu\text{m}.$

Optical Systems for Flow Parameter Measurement

9.1. Principles of Laser Doppler Velocimetry (LDV)

Laser Doppler velocimetry (LDV; also called laser interferometric anemometry) has been widely used over the last 40 years as an effective method for measurements in flows of very different origins. Due to the ability to perform measurements with no intervention in the studied flow by a material sensor, like in other measurement methods, the LDV technique is exploited in numerous applications – from aeronautics and turbomachinery to ophthalmology and other medical fields.

Of all the flow parameters the velocity vector $\vec{q}(u, v, w)$ is of main concern. If a laminar flow is investigated the steady-state velocity distribution is measured. A turbulent flow is a much more complicated situation characterized by a number of parameters. Turbulent flow velocity is a fluctuating vector, namely $u = \vec{u} + u'$, $v = \vec{v} + v'$; $w = \vec{w} + w'$, where $\vec{u}, \vec{v}, \vec{w}$ are the time-averaged values of *X*, *Y*, *Z*-components of velocity and u', v', w' are their fluctuations (randomly changing instantaneous values). As is well known, it is fluctuations that allow one to calculate the turbulent intensity (defined as $\varepsilon = \sqrt{(u')^2}/\vec{u}$, etc.) and to estimate the Reynolds stresses of the flow in terms of correlation functions $\vec{u'v'}, \vec{v'w'}, \vec{u'w'}$. Evidently all this requires a special approach which enables one to carry out very fast, numerous measurements.

Generally, LDV works as follows. Very small particles (tracers) are introduced (seeded) into a fluid. These particles should be small enough in order to follow the flow properly. The particles are illuminated by a laser beam and the scattered light parameters are measured by a remote detector yielding information about the



FIGURE 9.1 Fringe pattern in a probe volume.

particle velocity. It is postulated that the velocity of the particles at any chosen point in space represents the fluid velocity at that point.

The optical principle of measurement is demonstrated in Fig. 9.1. A laser beam is split initially into two parts which then cross each other in a probe volume (a small area around the point of measurements). Interference occurs in the probe volume and the spacing δ between interference fringes is governed by the angle, 2θ , between the two split beams:

$$\delta = \lambda/2\sin(\theta) \tag{9.1}$$

where λ is the wavelength of the laser beam. As a particle of velocity *u* (perpendicular to the direction of the fringes) travels through the measurement volume the scattered light intensity varies with frequency.

$$f = \frac{u}{\delta} \tag{9.2}$$

causing detector signal oscillations of the same frequency. By processing the detector signal its frequency f is found and then the velocity of the particle is calculated as follows:

$$u = \frac{\lambda}{2\sin(\theta)}f.$$
(9.3)

An optical arrangement for the realization of the above approach is illustrated in Fig. 9.2. A laser beam is divided by a beam splitter BS into two beams separated by a distance *l* and parallel to one another. These beams are collected by lens L_1 in a probe volume M which is located around the back focus of the lens. While a particle of velocity $\overline{q}(u, v)$ moves through the probe volume it scatters radiation in all directions, including the direction of the collecting optics (lens L_2 followed by diaphragm D and detector Ph, usually a photomultiplier or a photodiode). When the particle is approaching the maximum of the interference pattern the amount of



FIGURE 9.2 Optical arrangement for measurement of a single component of velocity.

light transmitted to the detector is increased. Conversely, if the particle approaches the minimum of the interference pattern the amount of radiation collected by the detector is reduced. As a result, the detector photocurrent is an oscillating function of time, as far as the transit time of the particle in the measurement volume is concerned. An example of the detector signal burst (a signal caused by a single particle) is shown in Fig. 9.3. As we see, the burst is a periodic function with variable amplitude. The amplitude variation results from the fact that the light intensity of the laser beam is not constant across the beam, but rather Gaussian (see Section 3.3):

$$I \cong I_0 \exp\left(-\frac{x^2 + y^2}{r_0^2}\right) \tag{9.4}$$



FIGURE 9.3 LDV signal burst.

where r_0 is the beam radius. Since two Gaussian beams interfere in the probe volume, the corresponding interference pattern is described as

$$I \cong I_0 \exp\left(-\frac{x^2 + y^2}{r_0^2}\right) \cos^2\left(\frac{\pi x}{\delta}\right).$$
(9.5)

Assuming that: (i) the particle velocity is perpendicular to the fringes (x = ut); (ii) radiation scattered by the particle is proportional to the intensity of light at the instantaneous location of the particle; and (iii) the detector is linear ($i_{det} = kI$), we get

$$i_{\text{det}} = C_{\text{S}}kI_0 \exp\left[-\frac{u^2(t-t_{\text{C}})^2 + y^2}{r_0^2}\right] \cos^2[\pi f(t-t_{\text{C}})]$$
(9.6)

where C_S is the cross-section of scattering of the particle and t_C is the time of arrival of the particle at the center of the probe volume (x = 0). Expression (9.6) describes the ideal signal burst shown in Fig. 9.3. Once the transit time, τ_S , between two adjacent fringes separated by a distance δ is measured, it can be immediately converted to frequency, $f = 1/\tau_S$, and then the velocity is calculated from Eq. (9.3).

The other values characteristic of arrangements like that of Fig. 9.2 and useful for the design of LDV systems are the size of the probe volume and the full number of fringes. Actually the measurement volume consists of two cones touching each other by their circular bases. The maximum diameter, d_m , of the volume with fringes is related to the focal length of lens L₁ (we denote it here as F'_1) and the divergence angle, ϑ , of the laser beam:

$$d_{\rm m} = 2\vartheta F_1' = \frac{2\lambda}{\pi w_0} F_1' \tag{9.7}$$

(w_0 is the laser waist radius, see Section 3.3). The length of the probe volume, l_m , depends on the angle 2θ between two beams (see Fig. 9.1) determined by the separation distance l after the beam splitter BS:

$$l_{\rm m} = \frac{d_{\rm m}}{2\tan(\theta)} = \frac{d_{\rm m}}{l}F_1^{\prime}.$$
(9.8)

Evidently the maximum number of fringes, N, in the probe volume can be found as

$$N = d_{\rm m}/\delta. \tag{9.9}$$

In the configuration shown in Fig. 9.2 the collection optics is positioned along the optical axis of the illumination system. In such a case (known as forward scattering architecture) the direct beams should be closed by a non-transparent stop with an opening which allows only the scattered light to come to the detector. Shown by the dotted lines in Fig. 9.2 is the back scattering arrangement: this includes an additional beam splitter which reflects the scattered light gathered by lens L_1 to lens L_3 followed again by an aperture D and a detector. The advantages of the second arrangement become evident in situations where measurements have to be performed in different areas of the studied flow. The forward scattering mode requires a realignment of the collecting optics any time the probe volume M is moved whereas the back scattering assembly remains unchanged.

In a highly turbulent flow the velocity vector of the seeded particle can be arbitrarily directed. Two particles with velocity components u and -u will cause the same burst (Eq. (9.6)) and therefore cannot be distinguished. To solve this problem (known as directional ambiguity) an additional element is introduced in the arrangement shown in Fig. 9.2. An acousto-optical modulator (AOM) (see detailed description in Section 7.2) is introduced in one of the beams incident on lens L₁. The AOM is aligned in such a manner that the first order diffracted beam emerging from the AOM is parallel to the second (undisturbed) beam leaving the beam splitter BS. Since the beam passing through the AOM is frequency shifted (say, by a value f_{ac}) with regard to the second one, a beat frequency occurs between the two beams and the fringe pattern in the probe volume is not steady, but moves in the direction normal to the fringes (direction OX). As a result, a particle moving in direction OX causes a signal burst with oscillating frequency $f - f_{ac}$ whereas a particle moving in the opposite direction creates a burst of frequency $f + f_{ac}$.

It should also be noted that the optical arrangement of Fig. 9.2 allows one to measure only one component (u) of the velocity vector. 2-D and 3-D measurement architectures are described in numerous publications devoted to the LDV technique (e.g., see references in Brown, 1986, Chapter 12). Some configurations are considered in Section 9.2.

Problems

9.1. Velocity measurements are carried out in a highly turbulent ($\varepsilon = 0.3$) transonic air flow ($U_{\text{max}} = 1$ Mach) using an LDV system capable of creating a probe volume with fringe spacing $\delta = 10 \,\mu$ m. The system includes an AOM allowing for a frequency shift of 50 MHz (to avoid ambiguity in the interpretation of captured signals).

- (a) What range of working frequencies (bandwidth) of the signal processing unit is required in order to investigate the statistics of the flow completely?
- (b) What happens if the AOM is limited to 40 MHz shifting?

9.2. What are the minimum and the maximum velocities which can be measured with the 1-D LDV system shown in Fig. 9.4 if the laser is operated at a wavelength $\lambda = 0.63 \mu m$ and has a waist diameter $2w_0 = 0.8 mm$, and the signal processor operates with frequencies up to 30 MHz?



FIGURE 9.4 Problem 9.2 – 1-D LDV system.

9.3. LDV with side scattering (off-axis) operation mode. Velocity measurements by an LDV system with large fringe spacing requires a small intersection angle θ which results in a very long measurement volume. To reduce the effective probe volume the side scattering mode is exploited. Assuming that collecting optics consists of two lenses (L₁ of $F'_1 = 250$ mm and L₂ of $F'_2 = 350$ mm) separated by 40 mm, a pinhole, and a photodetector (the other geometrical parameters are shown in Fig. 9.5), calculate the diameter of the pinhole *D* required for optimal measurement configuration.

[Note: The system operates with an argon laser ($\lambda = 0.514 \ \mu m$) and the probe volume size is 0.5 mm.]



FIGURE 9.5 Problem 9.3 – LDV with side scattering.

9.4. An LDV system is designed for the investigation of the velocity field, V, across a tube of diameter 2l = 20 cm having a transparent segment in the wall (see Fig. 9.6). The flow is laminar and its velocity obeys the equation

$$V = 5 + 10 \left[1 - \left(\frac{x}{l}\right)^2 \right]$$
m/s.



FIGURE 9.6 Problem 9.4 – LDV system for measurement of the velocity field in a tube.

The following elements are available for the system. (i) Laser: He–Ne; (ii) lenses: L₁, positive, $F'_1 = 100$ mm; L₂, negative, $F'_2 = -30$ mm; (iii) beam splitter with beam separation of 20 mm; (iv) detector assembly (with lens, pinhole, and photomultiplier).

- (a) How should one arrange the system so that it is capable of performing measurements at all points of the cross-section of the tube with a minimum of moving elements?
- (b) What changes in detecting signals are expected while the probe volume is moved from point A to B and C?

[Notes: (i) Due to constraints of the system mechanics the maximum distance between L_1 and L_2 cannot exceed 80 mm; (ii) the thicknesses of both lenses can be neglected.]

9.5. When the scale of turbulence in a flow is investigated a spatial correlation $\overline{u(A)u(B)}$ between velocities at different points along OZ is measured by the LDV system shown in Fig. 9.7. This includes an AOM made of TeO₂ ($V_S = 600 \text{ m/s}$) and activated by RF signals of frequency $f_{ac} = 30-50$ MHz. The AOM is aligned in such a manner that the zero order and the first order diffracted beams have almost the same intensity. If no RF is applied to the AOM only the zero-order beam (actually the initial radiation of the laser) is transferred through it and is split thereafter by beam splitter BS₁ and mirror M into two beams separated by 10 mm and concentrated by lens L₁ at point A in the flow. As the AOM is energized by RF power of frequency f_{ac} the first order diffracted beam arises in a direction different from the zero order and therefore the corresponding new pair of beams are concentrated by L₁ in another point, B. The location of B in the flow varies, as different frequencies are introduced in the AOM. The receiving optics, configured as a back scattering arrangement, consists of lens L₂, two pinholes, and two


FIGURE 9.7 Problem 9.5 – LDV system for measurement of spatial correlation of velocity.

photodiodes Ph A and Ph B, each one collecting signals of corresponding points A or B and transferring them to the signal processor where finally the correlation function is calculated. While B is moved along OZ the position of Ph B (and the pinhole) should be moved accordingly. The focal lengths of L₁ and L₂ are 500 mm and 250 mm, respectively, and the system works with a He–Ne laser with $\lambda = 0.63 \ \mu m$ and $2\theta = 2 \times 10^{-3}$.

- (a) What is the correlation length AB if $f_{ac} = 40$ MHz and 50 MHz?
- (b) How should one arrange the pinhole and Ph B in order to avoid cross-talk (influence of signals from A on B and vice versa)? Is it enough to put interference filters in front of each detector?
- (c) Assuming the reflectance of mirror M is 100%, what is the optimal ratio R/T (reflectance/transmittance) of the beam splitter exploited in the system?

9.2. Measurement of Velocity in 2-D and 3-D Flow Geometry

In many real situations a studied flow cannot be described in terms of 1-D geometry. In such cases the measurement of two or even three components of velocity becomes essential.

In order to measure two components, u and v, of velocity vector \vec{q} usually two interference patterns are created simultaneously in a probe volume. Figure 9.8 demonstrates one possible architecture of such a system. The light source is a laser generating two wavelengths, λ_1 and λ_2 (usually an argon laser with green (514 nm) and blue (488 nm) spectral lines). One of them is reflected by beam splitter BS₁,



FIGURE 9.8 Optical configuration of LDV system for measurement of two components of velocity.

filtered out by interference filter F_1 , and arranged by mirrors M_1 and M_2 and beam splitter BS₂ as two parallel beams incident on lens L_1 in the vertical plane. The second wavelength is transmitted by BS₁, filtered out by F₂, and split by BS₃ into two parallel beams striking lens L_1 in the horizontal plane. The lens focuses all four beams into the probe volume, M, of the flow where two sets of interference fringes are created: one is horizontal enabling one to measure the vertical component, *u*, of velocity and the second is vertical allowing for the measurement of the horizontal component, *v*. When a sampling particle moving with the flow crosses the probe volume it scatters simultaneously both wavelengths in all directions. Part of the scattered energy is collected by lens L_2 followed by filter F₃ (identical with F₁) and is transferred to detector Ph₁. Another part of the scattered energy is collected by lens L_3 and then passes through filter F₄ (identical to F₂) to the second detector, Ph₂. The signal of each detector is transferred to a separate processor where it is processed in a way similar to that described in Section 9.1 for a single velocity component. Hence, finally one obtains

$$u = f_1 \delta_1; \quad v = f_2 \delta_2 \tag{9.10}$$

where δ_1 and δ_2 are the fringe spacing at wavelengths λ_1 and λ_2 , respectively, and f_1 and f_2 are the oscillation frequencies of the corresponding detector bursts. The receiving optics shown in Fig. 9.8 (lenses L₂, L₃, filters F₃, F₄, and detectors Ph₁, Ph₂) is arranged in a side scattering (off-axis) operation mode (see details in Problem 9.3). This is not essential and in some cases it is more convenient to use the back scattering configuration (like that depicted by the dotted lines in Fig. 9.2).



FIGURE 9.9 Optical configuration of LDV system for measurement of three components of velocity.

Measurement of all three components of the velocity vector, u, v, and w, requires either three wavelengths or two wavelengths and two polarizations. An example is presented in Fig. 9.9 where a light beam emerging from a laser is divided by a polarizing beam splitter, BSP, into two polarization components, P and S, for both wavelengths λ_1 and λ_2 . The P component creates in a probe volume M two interference patterns using a 2-D optics channel while the S component creates in the same location M an additional (the third) interference pattern related to one of the working wavelengths. The optical axis of the S-component channel creates in the horizontal plane an angle α with the axis of the first channel. The fringes of the third pattern have to be arranged vertically. Obviously three detectors are operated simultaneously in the system, and the receiving optics of both channels can be arranged either in the side scattering mode (this case is demonstrated in the figure) or in the back scattering mode.

A particle traveling through the probe volume M scatters simultaneously radiation of both wavelengths and both polarizations. Part of the scattered light related to the P component is collected by lens L₃, filtered out by polarizer P₁, and split to detectors Ph₁ and Ph₂. Another part of the scattered light is collected by lens L₄ followed by polarizer P₂ (which transmits only the S component of radiation scattered by the particle) and proceeds further to detector Ph₃. The signal processor of this detector reveals the horizontal component v' of the velocity vector \vec{q} as it is projected on the plane XOY' constituting an angle α with the plane XOY. Since usually α is much smaller than 90° additional consideration of the vector components is involved.



FIGURE 9.10 (a) Geometry of velocity vector projection on two vertical planes and (b) relations between v, v', and w.

The 3-D geometry of vector projections on both vertical planes (perpendicular to the optical axis of both channels) is shown in Fig. 9.10a and Fig. 9.10b demonstrates the relation between two horizontal components, v, v', and the third component, w, of the velocity vector. It can be easily seen that the following relations exist between the velocity vector components:

$$w = v \times \tan \varphi; \quad \tan \varphi = \frac{k - \cos \alpha}{\sin a}; \quad k = \frac{v'}{v}.$$
 (9.11)

As we see, all three components are measured simultaneously. More details on LDV system configurations and signal processing techniques can be found in Brown (1986) and Durst (1982).

Problems

9.6. Flow parameters are measured with a 2-D LDV system operated at two wavelengths, $\lambda_1 = 0.55 \ \mu\text{m}$ and $\lambda_2 = 0.48 \ \mu\text{m}$. The system (see Fig. 9.11) includes two plano-convex cylindrical lenses, L₁ with a radius of 200 mm and thickness of 5 mm and L₂ with a radius of 100 mm and thickness of 7 mm, both made of glass with refractive index n = 1.5. Two beam splitters enable one to separate beams by 20 mm in the vertical and in the horizontal planes.

- (a) Find the location of the probe volume and the distance between the lenses.
- (b) Find the direction of the flow velocity if the measured frequencies of the detector signals are $f_1 = 0.5$ MHz and $f_2 = 0.3$ MHz.



FIGURE 9.11 Problem 9.6 – 2-D LDV system with cylindrical lenses.

(c) What is the maximum velocity the system is capable of measuring if the maximum frequency processed in each channel can be as high as 10 MHz?

9.7. A 2-D LDV system (depicted in Fig. 9.12), working with two wavelengths, $\lambda_1 = 515$ nm and $\lambda_2 = 430$ nm, includes lens L of radii $R_1 = -R_2 = 100$ mm and thickness t = 5 mm made of BK-7 glass ($n_1 = 1.519$; $n_2 = 1.523$), two beam splitters with beam separation $b_1 = b_2 = 20$ mm each, and receivers with two detectors operated in the back scattering mode. In the probe volume the velocity of flow is V = 10 m/s and the velocity vector is tilted 45° to the horizontal axis.

- (a) Calculate the frequencies measured at each channel.
- (b) Due to chromatic aberration of the lens the measurement points O_1 and O_2 in both channels do not coincide with one another. Calculate the distance O_1O_2 .
- (c) In order to perform measurements at the same (single) point in the flow it is decided to exploit the spherical aberration of the lens. Assuming that



FIGURE 9.12 Problem 9.7 – 2-D LDV system with two beam splitters and a single lens.

the lateral spherical aberration obeys the relation $\delta_s = 11.1 \times 10^{-3} r^2$, find how to change the beam separation in one of the channels.

9.8. A 3-D LDV system is operated with two wavelengths, $\lambda_1 = 0.488 \ \mu\text{m}$ and $\lambda_2 = 0.514 \ \mu\text{m}$, with P-polarization in two of the branches and wavelength $\lambda = 0.514 \ \mu\text{m}$ and S-polarization in the third branch. The angle between P and S branches is $\alpha = 25^{\circ}$. The system comprises three identical beam splitters, with beam separation $l_1 = l_2 = l_3 = 50 \ \text{mm}$, and two lenses of focal length 1 m each. RF frequencies measured in the first branch are $f_1 = 5.0 \ \text{MHz}, f_2 = 3.0 \ \text{MHz}$ and in the second branch $f_3 = 2.85 \ \text{MHz}$. Calculate the magnitude of the velocity in the probe volume of the flow.

9.3. Two-phase Flow and Principles of Particle Sizing

Two-phase flow is usually a mixture of a gas with liquid or solid particles or a liquid where solid particles or gas bubbles are present. In many practical applications the measurement of the velocity profile of a two-phase flow is accompanied by measurement of the statistics of particles with regard to their size. Numerous methods of particle sizing have been known for many years. Here we mention only those which are related to the LDV technique described above.

If a particle moving with a flow is much smaller than the fringe spacing δ in the probe volume the signal burst originating from the scattering of light by the particle has the shape presented in Fig. 9.3. However, if the particle size approaches δ or is even greater than the situation is different: the minima of the signal burst cannot approach zero, even if two interfering beams have equal intensities, since at any moment some portion of the particle is illuminated by light of the fringe maxima. As a result, the LDV burst becomes an oscillating function with two envelopes: the first related to the maxima and the second related to the minima. This situation is demonstrated in Fig. 9.13 where a large particle ($d > \delta$) is shown at three sequential moments when it moves through the fringes. The larger the particle the smaller the difference between the upper and lower envelopes of the signal burst (both envelopes are characterized by their amplitudes, I_{max} and I_{min} , shown in Fig. 9.13b).

Of course, the amount of radiation energy scattered by the particle and collected by the receiving optics is strongly dependent on the particle size, so that the amplitude I_{max} can be used as a measure of the particle diameter. The dependence $I_{\text{max}} = F(d)$ can be described by a square power law

$$I_{\max} = Cd^2 \tag{9.12}$$



FIGURE 9.13 (a) A large particle moving through a probe volume and (b) the corresponding signal burst.

and this simple formula remains valid over a wide range of particle sizes (from several micrometers to tenths of millimeters). The constant *C*, however, depends on the parameters of the measurement system (like laser power, detector sensitivity, collecting optics configuration, etc.) and also on the location of the particle trajectory inside the probe volume (see Problem 9.9). All this causes difficulties in exploiting Eq. (9.12) in practice. Naturally normalized values independent as much as possible of optical configuration would be much more convenient for practical applications. A method widely used is based on the measurement of visibility function, *V*, defined as the ratio of the AC to DC components of the signal burst generated by a particle. In terms of I_{max} and I_{min} shown in Fig. 9.13 the visibility function can be described as

$$V = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}}.$$
(9.13)

For a spherical particle traveling through an ideal fringe pattern the visibility function can be approximately expressed in terms of a Bessel function of the first order:

$$V = 2J_1(ka)/ka \tag{9.14}$$

where *a* is the radius of the particle and $k = 2\pi/\delta$. The corresponding graph is presented in Fig. 9.14. For any registered signal burst generated by the studied particle the values I_{max} and I_{min} are measured and visibility *V* is calculated from Eq. (9.13). Then, using Eq. (9.14) or the graph of Fig. 9.14, the corresponding value of the parameter $p = d/\delta$ is found and the particle diameter $d = p\delta$ is easily calculated if the fringe spacing δ is known. As can be seen from Fig. 9.14, the fringe spacing should be appropriately chosen in order to ensure that visibility is in the range 1.0 < V < 0.15, where *V* is a monotonic function of *p* and where



FIGURE 9.14 Visibility function for forward scattering mode (solid line) and for off-axis scattering mode (dotted line).

each measured value of V corresponds to a single possible value of the parameter p and therefore to a unique particle diameter d.

In reality the simple function of Eq. (9.14) is not always valid, especially if the off-axis configuration of an LDV system is exploited. The reason is the complexity of scattering phenomena. Indeed, scattering of radiation even by a particle of the simplest shape (spherical) is described by Mie formulas which are very complex and cumbersome (see the rigorous description in Born and Wolf, 1968). Mie's solution of the Maxwell equations predicts correctly the angular distribution of scattered radiation for spherical particles of any size and refractive index. Examples of such distributions, depicted in Fig. 9.15, demonstrate that the intensity of scattered light can vary significantly with observation angle. This is due



FIGURE 9.15 Angular diagrams of scattered radiation by (a) a small particle and (b) a large particle. The parameter $q = \pi d/\lambda$.

to interference of secondary waves generated by the particle illuminated by incident radiation. The larger the particle the more complex the interference pattern accompanying scattering of radiation.

The amount of scattered light absorbed by the receiving optics of an LDV system depends on both observation angle and collection angle, as well as on the refractive index of the particles. As a result, visibility also becomes dependent on these parameters and can vary noticeably, as can be seen from the dotted line shown in Fig. 9.14 (this curve and other cases are considered in detail in Bachalo *et al.*, 1980; a rigorous consideration taking into account the actual radiation field of scattered light when a moving particle is illuminated by two coherent light beams can be found in Durst, 1982).

Problems

9.9. A 1-D LDV system comprising a He–Ne laser with beam waist $w_0 = 0.2$ mm, a beam splitter giving 30 mm beam separation, and a lens of 200 mm focal distance is used for the measurement of particle size in a two-phase flow. The particles to be measured are in the range 20–100 µm. Choose an appropriate method of measurement and estimate the error resulting from the fact that the particle trajectory is perpendicular to the optical axis, but crosses the probe volume 1 mm to the side of the center point.

9.10. An LDV signal burst generated by a spherical particle of unknown size has two envelopes, an upper one and a lower one, and their maximum values are related to each other as 2:1. Assuming the LDV system is operated in the forward scattering mode and the fringe spacing is 15 μ m, find the size of the particle.

9.11. How should one choose the fringe spacing in an LDV system exploited for particle sizing based on visibility if the particle diameter in the flow can be as large as $50 \,\mu\text{m}$?

9.4. Solutions to Problems

9.1. (a) The maximum average velocity \overline{u}_{max} in the studied flow is 1 Mach, which corresponds to 330 m/s in air at normal conditions. Turbulence causes fluctuation changes as high as 30% of the average value: $\sqrt{(u')^2} = \varepsilon(\overline{u}) = 0.3 \times 330 = 99.9$ m/s. Therefore, the maximum instantaneous velocity which might occur in the measurement volume is as high as 430 m/s. From Eq. (9.2) it follows that the maximum detector signal oscillation frequency, with no frequency shift, would be

 $430/(10 \times 10^{-6}) = 43$ MHz. Taking into account the frequency shift of 50 MHz we draw the conclusion that for particles traveling through the probe volume in the direction of moving fringes the maximum measured frequency could be (50 - 43) = 7 MHz and for the particles traveling in the opposite direction (relatively to the fringes they "move" faster) the maximum measured frequency might achieve (50 + 43) = 93 MHz. Hence, finally, the range of working frequencies is from 7 MHz up to 93 MHz.

(b) If the available frequency shift is 40 MHz it might occur that some particles traveling in the direction opposite to the moving fringes will be interpreted as those moving slowly with the fringes. That is, the particles of velocity u = 40 MHz $\times 10^{-5}$ m = 400 m/s and of higher velocities (up to 430 m/s) moving in reality in the negative direction (against the fringes) will be interpreted as traveling in the positive direction (with the fringes). Therefore, the frequency shift of 40 MHz is unacceptable since the statistics of the flow will be treated incorrectly.

9.2. First one should find the size of the probe volume and the fringe spacing. As shown in Fig. 9.4, the focal length of the illumination lens is 700 mm which gives from Eq. (9.7)

$$d_{\rm m} = \frac{2 \times 0.63 \times 10^{-3}}{3.141 \times 0.8}$$
700 = 0.351 mm.

Since the beam separation is 20 mm the angle of intersection of two beams in the probe volume is $2\theta = 2 \times \tan^{-1}(10/700) = 1.64^{\circ}$. Then from Eq. (9.1) we get $\delta = 0.63/[2\sin(0.82^{\circ})] = 22 \,\mu$ m. The number of full fringes in the measurement volume is $N = \inf[351/22] = 15$. The maximum velocity which could be measured (actually the maximum component normal to the fringes) is found from Eq. (9.2): $u_{\text{max}} = 30 \times 10^6 \times 22 \times 10^{-6} = 660 \,\text{m/s}$. The minimum value of *u* is theoretically approaching zero. However, if the velocity vector of a particle is directed in such a way that it crosses the probe volume with no intersection of any fringe (like the vector \bar{q} shown in Fig. 9.16) then the tracer cannot be revealed by the system. The limiting direction is $\varphi = \tan^{-1}(22/351) = 3.59^{\circ}$. Therefore, the range $\pm 3.59^{\circ}$ is beyond the capability of the system.



FIGURE 9.16 Problem 9.2 – Limitation in measurement of a velocity vector direction.



FIGURE 9.17 Problem 9.3 – Geometry of probe volume.

9.3. The length of the probe volume is related to its maximum diameter, d_m , as

$$l_{\rm m} = \frac{d_{\rm m}}{\tan(\theta)} = \frac{0.5}{\tan(5^\circ)} = 5.7 \,\rm{mm}$$

Then, from the geometry of the probe volume (see Fig. 9.17) we can calculate the size of the segment AB which is imaged by the collecting optics to pinhole D:

AB =
$$d_{\rm m} \frac{\sin(85^\circ)}{\sin(180^\circ - 85^\circ - 45^\circ)} = 0.5 \frac{\sin(85^\circ)}{\sin(50^\circ)} = 0.65 \text{ mm}.$$

Using the paraxial approximation for collecting optics we find first the image of AB through L_1 and then through L_2 :

$$S_{1} = -400; \quad \frac{1}{S_{1}'} = \frac{1}{250} - \frac{1}{400}; \quad S_{1}' = 666.7 \text{ mm}; \quad V_{1} = -\frac{666.7}{400} = -1.67$$

$$S_{2} = 666.7 - 40 = 626.7 \text{ mm}; \quad \frac{1}{S_{2}'} = \frac{1}{350} + \frac{1}{626.7}; \quad S_{2}' = 224.6 \text{ mm};$$

$$V_{2} = \frac{224.6}{626.7} = 0.358$$

$$V_{\text{tot}} = V_{1}V_{2} = -0.598.$$

Therefore, $D = AB \times V_{tot} = 0.65 \times 0.598 = 0.389$ mm at a distance of 224.6 mm behind lens L₂.

9.4. (a) The system should be designed according to the back scattering configuration and the only moving element should be the negative lens L_2 (see Fig. 9.18). As the rays between L_1 and L_3 are parallel, changing the measuring point and corresponding movement of the negative lens do not affect the position of the pinhole in the receiving assembly.

(b) Consideration of point A. The velocity at this point is 5 m/s (since x = -1). To find the fringe spacing we take into account that $h_1 = 10$ mm and that the distance



FIGURE 9.18 Problem 9.4 – Relocation of the second lens in a 1-D LDV configuration.

between the two lenses in this position is a maximum and equal to 80 mm. This yields

$$h_2 = h_1 \frac{(100 - 80)}{100} 10 = 2 \text{ mm}; \quad S_2 = 20 \text{ mm};$$

 $\frac{1}{S'_2} = \frac{1}{-30} + \frac{1}{20}; \quad S'_2 = 60 \text{ mm}.$

Therefore, the intersection angle between two beams at point A is $\theta = \tan^{-1}(2/60) = 1.91^{\circ}$ and the fringe spacing (from Eq. (9.1)) is $\delta = 0.63/2$ sin $(1.91^{\circ}) = 9.45 \,\mu\text{m}$, which gives for the frequency of the detected signals $f_{\rm A} = 5/9.45 \times 10^{-6} = 0.53 \,\text{MHz}.$

Consideration of point B. The velocity at this point is 15 m/s (since x = 0). To move the intersection point from A to B (100 mm to the right) we have to move lens L₂ closer to L₁, i.e., to move it left by segment *z*. Therefore $S_2 = 20 + z$; $S'_2 = 60 + z + 100$, and we get the following equation with regard to *z*:

$$\frac{1}{160+z} - \frac{1}{20+z} = \frac{1}{-30}.$$

Solving this equation we find z = 5.4 mm. Then, as in the case of point A, we find the height of the side ray at L₂, $h_2 = 10(25.4/100) = 2.54$ mm, and calculate the new fringe spacing $\delta = 0.63/2(2.54/165.4) = 20.5 \mu$ m, and finally the signal frequency is $f_{\rm B} = (15/20.5) \times 10^6 = 0.73$ MHz.

Consideration of point C. The velocity at C is 5 m/s since x = 1. The intersection point moves right an additional 100 mm. Proceeding as in the previous case we get

$$S_2 = 20 + z;$$
 $S'_2 = 260 + z;$ $\frac{1}{260 + z} - \frac{1}{20 + z} = \frac{1}{-30};$ $z = 7 \text{ mm}.$

Then the height on lens L₂ is $h_2 = (27/100)10 = 2.7$ mm, the fringe spacing at point C is $\delta = 0.63/2(2.7/267) = 31.15 \,\mu$ m, and the signal frequency is $f_C = (5/31.15) \times 10^6 = 0.16$ MHz.

9.5. (a) The correlation length AB is dictated by the direction of the first order diffracted beam originating in the AOM. Using for the AOM the relations explained in Section 7.3 we define deviation of the first order from the zero order as $\Delta \alpha = (\lambda/V_S)f_{ac}$. Therefore, if the RF frequency activating the AOM is 40 MHz the corresponding deviation is $(0.63 \times 10^{-6})/3,600 \times 40 \times 10^{6} = 7 \times 10^{-3}$ and the distance $(AB)_1 = 7 \times 10^{-3} \times 500 \text{ mm} = 3.5 \text{ mm}$. In the second case, when the RF frequency is 50 MHz the correlation length $(AB)_2 = (0.63 \times 10^{-6})/3,600 \times 50 \times 10^{6} \times 0.5 = 4.375 \text{ mm}$.

(b) Points A and B are imaged by the illumination optics and receiving optics into the plane of the pinholes located in front of detectors Ph A and Ph B. Since the rays are parallel between L₁ and L₂, the optical magnification in the imaging is equal to the ratio of focal lengths: V = 250/500 = 0.5. The location of point A in the flow does not vary because the zero-order direction is constant. Consequently, detector Ph A and its pinhole should remain on the optical axis of L₂ while the measurements are carried out. To determine the size of the pinhole we first calculate the size of the probe volume. Using Eq. (9.7) we get $d_m = 2 \times 10^{-3} \times 500 = 1.0$ mm and the corresponding diameter of the pinhole $D_A = d_m V = 0.5$ mm. The size of the second pinhole, D_B , is also 0.5 mm, but its center should be moved aside from the optical axis of the receiving optics by $(AB)' = V \times (AB)$, which gives 1.75 mm and 2.19 mm for 40 MHz and 50 MHz, respectively.

To avoid the necessity of moving the second pinhole as f_{ac} varies, a method enabling one to distinguish between radiation coming from points A and B is required. Although radiation in these two points is of different wavelengths, this difference originating in the wavelength shift of the diffracted beam with regard to the incident beam of the AOM is very small and undistinguishable for interference filters.

(c) The optimal situation is achieved when the interfering beams have the same light intensity: in this case the interference pattern has the maximum achievable contrast of fringes. Assuming the light intensity after the AOM (in the first diffraction order) be I_d , we get for the two beams striking lens L_3 : $I_1 = I_d T$ and $I_2 = I_d(1 - T)T$ which gives the ratio $I_2/I_1 = 1 - T$. Therefore, the smaller the transmittance T of the beam splitter the better the contrast of the interference fringes in the probe volume. Practically, however, we cannot decrease T very much since radiation scattered by the particles moving through the probe volume should be powerful enough. As a compromise one should choose R/T = 80%/20% which yields $I_1 = 0.2I_d$; $I_2 = 0.16I_d$.

9.6. (a) Obviously the point of measurement (the probe volume) has to be located in a position where all four beams cross each other, i.e., the mutual focus of both lenses. Lens L₂ has optical power in the vertical plane and its focal length is found from Eq. (1.12) (keeping in mind that $r_1 = \infty$):

$$\frac{1}{f_2'} = \frac{n-1}{-r_2} = \frac{0.5}{100}; \quad f_2' = 200 \text{ mm.}$$

Since the back principal plane of L_2 is a tangent to the lens, one can state that the probe volume is located at a distance of 200 mm from the L_2 curved surface. In the horizontal plane only L_1 possesses optical power and the second lens acts as a parallel slab of glass. Calculating the focal length of L_1 again from Eq. (1.12) gives

$$\frac{1}{f_1'} = \frac{0.5}{200}; \quad f_2' = 400 \text{ mm}$$

and taking into account that the parallel slab of thickness $t_2 = 7$ mm (the second lens) causes additional displacement $\Delta = t_2(n-1)/n = t_2/3$ of the point of intersection of the beams leaving L₁ (see Problems 1.6), we get from the geometry of rays (shown in Fig. 9.19) for the distance *d* between two lenses

$$d = f_1' + \frac{t_2}{3} - f_2' - t_2 = 400 - 200 - 14/3 = 195.3 \text{ mm.}$$

(b) Using Eq. (9.1) we calculate the fringe spacing in each channel. For the channel where lens L_1 generates fringes (wavelength $\lambda_1 = 0.55 \ \mu$ m) we have

$$\delta_1 = \frac{\lambda_1}{2\sin\theta_1}; \quad \theta_1 = \arctan(10/400) = 1.43^\circ; \quad \delta_1 = \frac{0.55}{2\sin(1.43^\circ)} = 11.0 \,\mu\text{m}$$



FIGURE 9.19 Problem 9.6 – Geometry of rays in the horizontal plane.

and for the second channel

$$\delta_2 = \frac{\lambda_2}{2\sin\theta_2}; \quad \theta_2 = \arctan(10/200) = 2.855^\circ; \quad \delta_2 = \frac{0.48}{2\sin(2.855^\circ)} = 4.8 \,\mu\text{m}.$$

Therefore, the measured components of velocity, u and v, are as follows: $u = \delta_1 f_1 = 11 \,\mu\text{m} \times 0.5 \,\text{MHz} = 5.5 \,\text{m/s}$; $v = \delta_2 f_2 = 4.8 \,\mu\text{m} \times 0.3 \,\text{MHz} = 1.44 \,\text{m/s}$ and the direction of the vector \overline{q} is defined by the angle $\varphi = \arctan(v/u) = \arctan(1.44/5.5) = 14.7^{\circ}$.

(c) The maximum measured values at each channel limited by the processing module are calculated in a similar way: $u_{\text{max}} = \delta_1 f_{\text{max}} = 11 \ \mu\text{m} \times 10 \ \text{MHz} = 110 \ \text{m/s}$; $v_{\text{max}} = \delta_2 f_{\text{max}} = 4.8 \ \mu\text{m} \times 10 \ \text{MHz} = 48 \ \text{m/s}$, which gives the maximum absolute value of velocity as $q = \sqrt{(110)^2 + (48)^2} = 120 \ \text{m/s}$.

9.7. (a) Using Eq. (1.12) we first calculate the focal length of lens L at two given wavelengths:

$$\frac{1}{f_1'} = (1 - 1.519)\frac{2}{100} - \frac{5(1.519 - 1)^2}{100^2 \times 1.519} = 0.01029; \quad f_1' = 97.17 \text{ mm}$$
$$\frac{1}{f_2'} = (1 - 1.523)\frac{2}{100} - \frac{5(1.523 - 1)^2}{100^2 \times 1.523} = 0.01037; \quad f_2' = 96.43 \text{ mm}$$

and then find the convergence angle between two beams at each channel and the corresponding fringe spacing, as per Eq. (9.1):

$$\tan \theta_1 = \frac{10}{97.17}; \quad \theta_1 = 5.88^\circ; \quad \delta_1 = \frac{0.515}{2\sin(5.88^\circ)} = 2.52 \,\mu\text{m}$$
$$\tan \theta_2 = \frac{10}{96.43}; \quad \theta_2 = 5.92^\circ; \quad \delta_2 = \frac{0.43}{2\sin(5.92^\circ)} = 2.08 \,\mu\text{m}.$$

This enables one to calculate the measured frequencies in both channels:

$$f_{ac1} = \frac{u}{\delta_1} = \frac{10\cos 45^\circ}{2.52 \times 10^{-6}} = 2.81 \text{ MHz};$$

$$f_{ac2} = \frac{v}{\delta_2} = \frac{10\sin 45^\circ}{2.08 \times 10^{-6}} = 3.40 \text{ MHz}.$$

(b) The distance O_1O_2 is equal to the difference between the two focal lengths corresponding to the two wavelengths: $O_1O_2 = 97.17 - 96.43 = 0.74$ mm.

(c) Due to spherical aberration of the lens the intersection of the beams in each channel occurs not in the focus but in the point located closer to the lens by the segment equal to the lateral spherical aberration. As the beams strike the lens at a distance r = 10 mm from the optical axis the corresponding spherical aberration is $\delta'_{\text{Sph}} = 11.1 \times 10^{-3} \times 10^2 = 1.11$ mm. Hence, point O₁ is located

at a distance 97. 17 - 1. 11 = 96.06 mm from the lens and point O₂ at a distance 96.06 - 0.74 = 95.32 mm. In order to cause all four beams to cross each other in a single point it is necessary to reduce the spherical aberration of the second channel by 0.74 mm, i.e., the spherical aberration in this channel should be 1. 11 - 0.74 = 0.37 mm. This will occur if each beam is distant from the optical axis by $r = \sqrt{0.37/(11.1 \times 10^{-3})} = 5.77$ mm. Therefore, it is necessary to change the beam separation of the beam splitter BS₂ to 11.54 mm (instead of 20 mm).

9.8. Using Eq. (9.1) we calculate the fringe spacing in the first channel for the wavelength λ_1 :

$$\delta_1 = \frac{\lambda_1}{2\sin\theta_1} = \frac{0.488}{2\sin[0.5\arctan(50/1000)]} = 9.77\,\mu\text{m}$$

and then the fringe spacing for the wavelength λ_2 in the second and third channels:

$$\delta_2 = \delta_3 = \frac{0.514}{2\sin[0.5\arctan(50/1000)]} = 10.3 \,\mu\text{m}.$$

This yields the following values for the measured components u, v, v' in all three channels: $u = \delta_1 f_1 = 9.77 \times 5 = 48.9$ m/s; $v = \delta_2 f_2 = 10.3 \times 3 = 30.9$ m/s; $v' = \delta_3 f_3 = 10.3 \times 2.85 = 29.36$ m/s. Furthermore, using Eq. (9.11) we get

$$k = \frac{v'}{v} = \frac{29.36}{30.9} = 0.95; \quad \cos \alpha = \cos 25^\circ = 0.906; \quad \sin \alpha = 0.4226$$
$$0.95 - 0.906$$

 $\tan \varphi = \frac{0.25 - 0.200}{0.4226} = 0.104; \quad w = v \tan \varphi = 30.9 \times 0.104 = 3.21 \text{ m/s}$ and finally $q = \sqrt{u^2 + v^2 + w^2} = \sqrt{48.9^2 + 30.9^2 + 3.21^2} = 57.93 \text{ m/s}.$

9.9. First we check if the method based on visibility can be applied in this problem. To do this we calculate the fringe spacing δ as described in Eq. (9.1):

$$\tan \theta = \frac{l/2}{f'} = \frac{15}{200} = 0.075; \quad \theta = 4.29^{\circ}; \quad \delta = \frac{0.63}{2\sin(4.29^{\circ})} = 4.2 \,\mu\text{m}.$$

Hence, for the particles to be measured the ratio $p = d/\delta$ varies from $p_1 = 20/4$. 2 = 4.76 to $p_2 = 100/4$. 2 = 23. 8. Both values are much greater than the value p_{\min} corresponding to the first zero point of Eq. (9.14) ($(ka)_{\min}^{(1)} = 1.22\pi$; $p_{\min} = d_{\min}/\delta = 1.22$), so that visibility cannot be used in this case and therefore the measurement should be based on Eq. (9.12).

To estimate the error caused by the location of the particle trajectory in the probe volume one should take into account that I_{max} is proportional to the light intensity at the axial point at which the trajectory crosses the optical axis. In our case this point is removed 1 mm from the center of the probe volume, or, in terms of radial displacement Δ_r in each of two interfering beams, it is equivalent to

 $\Delta_r = 1 \times \sin \theta = 0.075$ mm. The intensity of each laser beam in radial coordinate Δ_r is governed by the Gaussian function (see Eqs. (3.1) and (3.2)):

$$I = I_0 \exp\left(-\frac{2\Delta_r^2}{w^2}\right) = I_0 \exp\left[-8\left(\frac{\Delta_r}{d_m}\right)^2\right]$$
(A)

where we also take into account that $w = d_m/2$ with the probe volume size d_m defined as in Eq. (9.7): $d_m = (2 \times 0.63 \times 200)/0.2\pi = 0.4$ mm. By substituting the values of Δ_r and d_m in Eq. (A) and keeping in mind that the interference maximum is four times higher than the intensity of each interference beam we draw the conclusion that I_{max} for the trajectory which passes through the probe volume center is $4I_0$ and for the trajectory passing 1 mm to the side $I_{max} = 4I_{\Delta_r}$, and therefore the ratio of both values is $I_{\Delta_r}/I_0 = \exp[-8(0.075/0.4)^2] = 0.755$. Therefore, using Eq. (9.12) with the reduced intensity I_{Δ_r} yields the reduction in the calculated diameter $d: d_{\Delta_r}/d = \sqrt{0.755} = 0.869$ which means a reduction of 13% (the error) of the measured size of the particle.

9.10. We rewrite the visibility definition (Eq. (9.13)) dividing the numerator and denominator by I_{min} :

$$V = \frac{(I_{\text{max}}/I_{\text{min}}) - 1}{(I_{\text{max}}/I_{\text{min}}) + 1} = \frac{2 - 1}{2 + 1} = 0.333.$$

For this value we find from the graph (solid line) of Fig. 9.14 that p = 0.87. (Of course, it is also possible to solve by trial and error the nonlinear Eq. (9.14) – it yields the same result.) Therefore, $d = p\delta = 0.87 \times 15 = 13.1 \,\mu\text{m}$.

9.11. The visibility approach can be exploited as long as V is greater than 0.15. As is evident from Fig. 9.14, the corresponding value of $p = d/\delta$ should be less than 1.05 and therefore the fringe spacing should be $\delta = d_{\text{max}}/p_{\text{max}} = 50/1.05 = 47.6 \,\mu\text{m}$ at least.

Chapter 10

Color and its Measurement

10.1. Color Sensation, Color Coordinates, and Photometric Units

Color Vision

Color has not only a physical meaning: it is a combination of physical effects and the physiology of human sensation. Seeing is a physiological process originating in photochemical reactions in two kinds of cells, rods and cones, present in a human eye retina. The optics of the eye creates an image on the retina, simultaneously in all wavelengths incident on the eye pupil. The rod cells comprise a photopigment, rhodopsin, which is sensitive only to light intensity, regardless of its spectral composition, and for this reason the rods do not detect color. The cone cells can be divided into three groups, each one having a different photopigment: erythrolabe, chlorolabe, and cyanolabe. The first has maximum absorptivity in the wavelengths of the red part of the spectrum and the second and third have maximum absorptivity in the green and the blue parts of the visual spectrum, respectively. Therefore, the cone cells are responsible for color sensation.

The cones and rods from different parts of the retina are not fully identical, but vary in their morphological structure. The concentration of cones and rods also varies. As a result, for instance, the sensation of the central part of the retina differs from that of the periphery. Despite these differences it is commonly accepted that color sensation is characterized by three spectral curves, regardless of the location of the photoreceptor in the retina. The curves are shown in Fig. 10.1 (these data are presented in Buchsbaum (1981) and also in another form in Boynton (1972)). It is



FIGURE 10.1 Relative spectral sensitivity of three kinds of photoreceptors: 1, $R(\lambda)$, cones with erythrolabe; 2, $G(\lambda)$, cones with chlorolabe; 3, $B(\lambda)$, cones with cyanolabe.

important to notice that any wavelength as well as any combination of wavelengths from the visual range (0.4–0.7 μ m) cause a photoreaction of all three kinds of cells, but with different "strength." Thus, the sensation of color originates from a combined reaction of the three kinds of photoreceptors (cones) present in the retina.

Three-stimulus Generalization

Mathematically the above physical and physiological description of color can be expressed in the following manner. Any light source S with spectral radiometric flux $P(\lambda)$ generates three main stimuli in a human eye: red (*r*), green (*g*), and blue (*b*). The relative "strength" of each can be estimated by the integrals

$$r = \int P(\lambda)R(\lambda) \,\mathrm{d}\lambda; \quad g = \int P(\lambda)G(\lambda) \,\mathrm{d}\lambda; \quad b = \int P(\lambda)B(\lambda) \,\mathrm{d}\lambda \tag{10.1}$$

where $R(\lambda)$, $G(\lambda)$, and $B(\lambda)$ are the spectral sensitivity of the three pigments of the eye shown in Fig. 10.1 (we term them natural primaries). Perception of all three stimuli creates the feeling of color, Q, of the light source S. This color can be represented by its vector

$$\overline{Q} = r\overline{R} + g\overline{G} + b\overline{B} \tag{10.2}$$

and by a corresponding point Q(r, g, b) in the 3-D color space (see Fig. 10.2a). Hence, r, g, and b are considered as color coordinates of the color Q and therefore of the light source S.



FIGURE 10.2 (a) 3-D color space and (b) its 2-D presentation in a single plane.

If two colors, Q_1 and Q_2 , are mixed together then

$$\overline{Q}_1 + \overline{Q}_2 = (r_1 + r_2)\overline{R} + (g_1 + g_2)\overline{G} + (b_1 + b_2)\overline{B}.$$
(10.3)

The tristimulus theory establishes that any color can be generated by taking three basic colors (called primaries) in an appropriate proportion. If the basic colors are chosen as three terms of Eq. (10.2) based on the integrals of Eq. (10.1) then the primaries are the natural red, green, and blue colors. However, this is not the only choice and not the most convenient choice in some situations (e.g., it turns out that in creating some colors the natural primaries have to be not only added but also subtracted from one another). For this reason another set of primaries was suggested and adopted by the International Commission on Illumination (CIE). These primaries are based on the standardized color mixture curves, sometimes termed color matching functions (Inglis, 1993). The curves, shown in Fig. 10.3,



FIGURE 10.3 Color mixture curves (standardized by the CIE).

are not physically realized in any sensor, but constitute a convenient computational tool.

In terms of color mixture curves the tristimulus values related to the light source S are as follows:

$$X = \operatorname{const} \int_{0.38}^{0.77} P(\lambda)\overline{x}(\lambda) d\lambda; \quad Y = \operatorname{const} \int_{0.38}^{0.77} P(\lambda)\overline{y}(\lambda) d\lambda; \quad Z = \operatorname{const} \int_{0.38}^{0.77} P(\lambda)\overline{z}(\lambda) d\lambda$$
(10.4)

and they are usually converted to the normalized color coordinates x, y, and z:

$$x = \frac{X}{X + Y + Z}; \quad y = \frac{Y}{X + Y + Z}; \quad z = \frac{Z}{X + Y + Z}.$$
 (10.5)

Evidently x + y + z = 1, so that only two coordinates, say x and y, are enough to characterize the color of an object. The corresponding x, y-diagram is shown in Fig. 10.4. The "gray" color, or no color at all, is defined as a situation when all three normalized color coordinates are equal: X = Y = Z. In terms of x, y, and z this means:

$$x_g = y_g = z_g = 0.33. \tag{10.6}$$

This case is related to point O in Fig. 10.4. All gray levels, from black to white, are represented in this point.



FIGURE 10.4 *x*, *y*-diagram of colors (the CIE chromaticity diagram).

$$x_1 = 0.33 - x_2; \quad y_1 = 0.33 - y_2; \quad z_1 = 0.33 - z_2.$$
 (10.7)

Consideration of white (or gray) color also has an important additional meaning. Each color can also be considered as a mix of white and some pure color. These pure colors could be related to monochromatic light, as shown by the color curve (1) of Fig. 10.4. This graph indicates some limits on where the points representing color could be. Moving from point O in some chosen direction towards the color curve (1) means a transition from the colorless (gray) situation to a pure color characterized by a specific wavelength. More details can be found in Wiszecki and Stiles (1982) and Inglis (1993).

Different Coordinate Systems

From a physical point of view it is preferable to use color coordinates which have physical meaning and not just physiologically originated parameters like red, green, and blue primaries. One can create another type of color coordinates, separating the brightness (or light intensity), the relative fraction of white (or saturation), and a color itself (called hue). There exist different ways to define H, S, I (hue, saturation, intensity). We consider one of them (see Fig. 10.5) where hue is defined as an angle, φ , in the plane ABC of Fig. 10.5b; saturation is defined as an angle, ϑ , between the vector \overline{Q} and the line OO₁ perpendicular to ABC and passing through the points of gray color; and intensity is defined as a distance along OO₁ from the origin O to the plane ABC passing through point Q. The relations between H, S, I and r, g, b are as follows:

$$H = 120^{\circ} + \arctan\left[\frac{\sqrt{3}(b-g)}{2r-b-g}\right]; \quad S = \arctan\left[\frac{\sqrt{3}(b-g)^{2} + (2r-b-g)^{2}}{\sqrt{2}(r+g+b)}\right];$$
$$I = \frac{r+g+b}{\sqrt{3}}.$$
(10.8)

There are numerous examples where the use of Q(H, S, I) with color coordinates H, S, I from Eq. (10.8) is preferable over a description of color in terms of r, g, b. It should also be mentioned here that r, g, b values in Eq. (10.8) could be replaced by X, Y, Z from Eq. (10.4) or by any other primaries if they are properly defined (see, for example, the primaries described in Section 10.2 and accepted for video devices).



FIGURE 10.5 (a) *H*, *S*, *I* color coordinate system and (b) detail of plane ABC.

Photometric Units vs. Radiometric Units

Special features of the human eye made it necessary to develop a special system of units (photometric units) in order to characterize adequately the light-related values and to take into account the spectral sensitivity of the eye. Of all the photometric units the two most frequently used are lumen (lm), which is the unit of the energy flux, and lux (lx), which is the unit of illumination generated by a flux of one lumen incident on an area of 1 m^2 ($1 \text{ lx} = 1 \text{ lm/m}^2$).

The relations between the photometric units and the radiometric (standard) units are based on the spectral sensitivity of the human eye shown in Fig. 10.6 (this is related to the rodopsin pigment present in the rod cells of the retina). It is commonly known that a normal human eye achieves maximum sensitivity at 0.555 μ m



wavelength and the full range of wavelengths where the sensitivity of the eye differs from zero is from 380 nm to 770 nm. Monochromatic light flux of 1 W power at $\lambda = 0.555 \,\mu$ m is equivalent to 683 lm. Luminous efficacy, $K(\lambda)$, is the function which establishes the relation between the photometric flux (F_V) and radiometric flux (F_E) and it is measured in lm/W:

$$K(\lambda) = F_{\rm V}(\lambda)/F_{\rm E}(\lambda). \tag{10.9}$$

Luminous efficiency, $V(\lambda)$, is a dimensionless function defined as a normalized luminous efficacy:

$$V(\lambda) = K(\lambda)/K(0.555) = K(\lambda)/683.$$
(10.10)

It is $V(\lambda)$ that describes the relative spectral response of the human eye (shown in Fig. 10.6). Obviously, 1 W of red light causes less of a visual sensation than 1 W of yellow-green radiation.

Luminous efficacy *K* and luminous efficiency *V* can be calculated over any chosen spectral interval. For instance, a source of light illuminating radiation power of $P(\lambda)$ in the interval (λ_1, λ_2) can be characterized by the total efficacy, *K*, expressed by the formula:

$$K = \int_{0.38}^{0.77} F_{\rm V}(\lambda) \mathrm{d}\lambda / \int_{\lambda_1}^{\lambda_2} P(\lambda) \,\mathrm{d}\lambda = 683 \int_{0.38}^{0.77} P(\lambda) V(\lambda) \,\mathrm{d}\lambda / \int_{\lambda_1}^{\lambda_2} P(\lambda) \,\mathrm{d}\lambda. \quad (10.11)$$

Problems

10.1. Find the energy flux (in radiometric units) corresponding to a monochromatic luminous flux of 100 lumen at $\lambda = 0.67 \ \mu m$ and to a flux of 50 lumen at $\lambda = 0.5 \ \mu m$.

10.2. What is the luminous efficacy and the luminous efficiency of a black body at temperature T = 2,000 K exploited as a light source for the visible range?

10.3. If a minimum illumination level of a CCD is declared as 0.3 lx with a F/# 1.2 lens and saturation is achieved when the power density is as high as $8.4 \,\mu$ W/cm², what is the actual dynamic range of this sensor?

10.4. The color of an object is characterized by normalized color coordinates x = 0.2; y = 0.15. Find the complementary color for this object.

10.5. Find the color coordinates H, S, I for an object with normalized coordinates x = 0.2 and y = 0.5.

10.2. Color Detection and Measurement

Configuration of Color Detectors

Since no photosensitive material has a spectral sensitivity identical to that of a human eye, any electro-optical sensor for color detection should comprise not only a detector sensitive in the visible range, but also a filter (actually a set of filters) allowing for spectral correction of sensitivity. In most color detection devices a silicon detector is exploited, with spectral sensitivity covering both the visible and near-IR ranges (see Fig. 10.7, solid line). An IR cut-off filter, which is usually added to a black and white CCD detector, makes the spectral properties of the detector to be more similar to that of the eye (compare the solid and dashed lines in Fig. 10.7), but this is definitely not enough if correct color measurements is necessary. According to the discussion in Section 10.1, three different detectors are required for color measurements, each one for a separate basic component of color (red, green, and blue). To each detector an appropriate filter is attached in order to provide spectral correction of the detector (silicon) to red, green, or blue sensitivity function of the eye. Denoting the spectral response of the detectors as $D(\lambda)$ and the transmittance of the correction filters in red, green, and blue channels as $T_{\rm R}(\lambda)$, $T_{\rm G}(\lambda)$, and $T_{\rm B}(\lambda)$, we can define the detector primaries as

$$R_{\rm D}(\lambda) = D(\lambda)T_{\rm R}(\lambda); \quad G_{\rm D}(\lambda) = D(\lambda)T_{\rm G}(\lambda); \quad B_{\rm D}(\lambda) = D(\lambda)T_{\rm B}(\lambda).$$
 (10.12)

IR Filter transmittance, %





FIGURE 10.7 Spectral sensitivity (quantum efficiency) of a silicon detector without IR cut-off filter (1) and transmittance of the IR cut-off filter (2).

Then the color coordinates of a light source S, with spectral density distribution $P(\lambda)$ are

$$i_{\rm R} = \text{const} \int_{\lambda_1}^{\lambda_2} P(\lambda) R_{\rm D}(\lambda) \, \mathrm{d}\lambda; \quad i_{\rm G} = \text{const} \int_{\lambda_1}^{\lambda_2} P(\lambda) G_{\rm D}(\lambda) \, \mathrm{d}\lambda;$$
$$i_{\rm B} = \text{const} \int_{\lambda_1}^{\lambda_2} P(\lambda) B_{\rm D}(\lambda) \, \mathrm{d}\lambda \tag{10.13}$$

and for the normalized color coordinates we have

$$x_{\rm D} = \frac{i_{\rm R}}{i_{\rm R} + i_{\rm G} + i_{\rm B}}; \ y_{\rm D} = \frac{i_{\rm G}}{i_{\rm R} + i_{\rm G} + i_{\rm B}}; \ z_{\rm D} = \frac{i_{\rm B}}{i_{\rm R} + i_{\rm G} + i_{\rm B}}.$$
 (10.14)

For video cameras the primaries of Eq. (10.12) are standardized by an international committee (NTSC). The corresponding functions, presented in Fig. 10.8, provide compatibility with the CIE chromaticity diagram and with video display devices, the latter having primaries defined as brightness of three components, $B_{\rm R}(\lambda)$, $B_{\rm G}(\lambda)$, $B_{\rm B}(\lambda)$, at equal input electrical signals.

It should be mentioned here that the negative sections of the camera primary functions cannot be realized optically, but they can be achieved by camera electronics where the signals are processed prior to output. In many practical cases, however, the negative parts of the primaries shown in Fig. 10.8 are not taken into consideration (just replaced by zero).



FIGURE 10.8 Primaries of CCD video cameras (NTSC standard): 1, $R_D(\lambda)$, red channel function; 2, $G_D(\lambda)$, green channel function; 3, $B_D(\lambda)$, blue channel function.



FIGURE 10.9 Schematic of a three-CCD color sensor.

Area detectors for color imaging can be arranged in two configurations:

- (a) Three spatially separated black and white detectors are combined with an appropriate set of filters enabling transmittance of R, G, and B components to different sensors (a three-CCD camera is arranged in this manner, see Fig. 10.9). The beam splitter BS₁ in Fig. 10.9 reflects red light onto sensor D₁ and transmits the rest (green and blue), whereas BS₂ reflects green light to sensor D₂ and transmits the rest (blue) to sensor D₃.
- (b) A configuration is used with a single black and white CCD detector to which a mosaic of filters is attached, each filter being the size of a single pixel and positioned accordingly. Figure 10.10 demonstrates two kinds of mosaic (other kinds of mosaic are also used in practice). Since "red," "green," and "blue" pixels become spatially separated, an actual unit of resolution is at least twice that of a black and white CCD. Hence, a color CCD with a mosaic of filters is characterized by a degradation of spatial resolution and this is a noticeable shortcoming of single-CCD color sensors



FIGURE 10.10 Arrangement of a mosaic of filters in a color CCD sensor.

compared to three-CCD devices. However, the latter are more complex and correspondingly more expensive.

Configuration of Output Signals

CCD sensors are usually followed by electronic circuitry where the signals of red, green, and blue are arranged either in three separate channels or combined in a single composite video signal when each video line comprises three sequential sections corresponding to red, green, and blue pixels of that line.

No matter how the pixel signals are arranged, they undergo an additional transformation before output. To understand this transformation we should keep in mind that, in general, color video signals are used in two ways: (i) creation of color images on a video monitor display where they are visually observed; and (ii) digital processing and measurement of color image parameters (like color coordinates, light intensity, etc.). In the first case the display features and operation functions should be taken into consideration. Display screens are made of three layers of phosphors (materials emitting light) with well-specified display primaries (mentioned earlier) and the corresponding color coordinates related to the CIE chromaticity diagram (points S_R, S_G, and S_B indicated in Fig. 10.4). Since 1982 these coordinates have been standardized as follows (see detailed description in Inglis, 1993): red primary, x = 0.635, y = 0.34; green primary, x = 0.305, y = 0.595; blue primary, x = 0.155, y = 0.07. Electrical signals coming to the monitor are converted into electronic beams of the corresponding intensities which strike the phosphors which are energized accordingly and emit a mixture of their primary colors in a proportion governed by the video signal at the display input. Besides this, the brightness of the display screen, $B_{\rm S}$, is a non-linear function of the input electrical signal. This function is usually expressed as a power γ :

$$B_{\rm S} = \text{const} \times i^{\gamma} \tag{10.15}$$

where $\gamma = 2.2$ in most practical cases. As a result, the relation between the color primaries displayed by the screen differs from that received at the device input. To avoid this effect a gamma-correction procedure is performed by the CCD electronic circuitry where all pixel signals undergo transformation of $(1/\gamma)$ power before output.

In reality color video signals are arranged according to one of three internationally accepted standards and CCD electronics organizes the pixel signals according to a specified standard. A description of these standards is beyond the scope of this book; details can be found in Inglis (1993) and other books on video engineering.

Another issue which should be mentioned here is the white balance of color detectors. Since red, green, and blue channels have separate electronics, the amplification factors $A_{\rm R}$, $A_{\rm G}$, and $A_{\rm B}$ can be different, which might affect the relative

fraction of color components in the output signals. The procedure of white balancing is just on equalization of red, green, and blue signals (i_R, i_G, i_B) while the white (colorless) background is imaged on the detector. This procedure is especially important if an image is not only observed on a display but also processed with the aim of color measurement.

Problems

10.6. Which color will be displayed on the screen of a video monitor if the signals at the input of red, green, and blue channels are related as 0.8:0.3:1.0? [Note: The screen is made of phosphors fulfilling the requirements of SMPTE standard C (1982).]

10.7. A scene background illuminated by a narrow-band LED of 595 nm wavelength is imaged on a color CCD camera which has variable gamma correction and primaries conforming to the requirements of NTSC standard. The white balance of the camera is performed with an illumination source of color temperature 3200 K. Which color of the scene will be measured by the camera? (Find the *x*, *y* color coordinates and calculate the corresponding value of hue.)

10.3. Solutions to Problems

10.1. Using the definition of luminous efficacy and luminous efficiency from Eqs. (10.9) and (10.10), we have for the radiometric flux, F_E : $F_E = F_V/K(\lambda) = F_V/[683 \times V(\lambda)]$. The value of V can be found from the graph of Fig. 10.6. At a wavelength of 0.67 μ m this value is 0.03 and for $\lambda = 0.5 \mu$ m we have V = 0.32. This gives

$$(F_{\rm E})_1 = \frac{100}{683 \times 0.03} = 4.88 \text{ W}; \quad (F_{\rm E})_2 = \frac{50}{683 \times 0.32} = 0.229 \text{ W}.$$

10.2. We use the definition of luminous efficacy for the wide-band light source from Eq. (10.11) which in our case (radiation of a black body) yields

$$K = 683 \frac{\int_{0.38}^{0.77} e_{\rm B}(\lambda, T) \times V(\lambda) \,\mathrm{d}\lambda}{\int_{0}^{\infty} e_{\rm B}(\lambda, T) \,\mathrm{d}\lambda} = 683 \frac{I_1}{I_2} \tag{A}$$

where we denote by $e_B(\lambda, T)$ the spectral radiation of the black body, referred to in Section 6.1 and in Appendix 3. Of the two integrals, I_1 and I_2 , the latter is governed by the Stefan–Boltzmann law (see Chapter 6): $I_2 = \sigma T^4$. We calculate the former integral numerically, by dividing the visual range into 18 spectral intervals of $\Delta \lambda = 0.02 \ \mu m$ width each and exploiting the table and notation of Appendix 3. In doing this we first rewrite Eq. (A) as follows:

10

$$K = 683 \frac{\Delta \lambda \sum_{i=1}^{18} a_i V_i}{\sigma T^4} 10^{-5} \sigma T^5 = 683 \times T \times 10^{-5} \Delta \lambda \sum_{i=1}^{18} a_i V_i$$
(B)

where a_i are the values from the second column of Appendix 3 calculated for each (λ_i, T) using interpolation and V_i are the corresponding values of the spectral sensitivity of the eye taken from Fig. 10.6. After calculation we finally get $K = 683 \times 0.02 \times 10^{-5} \times 2,000 \times 6.196 = 1.69$ lm/W. With this value one can find the total luminous efficiency of the light source:

$$\eta = \frac{K}{683}\% = \frac{1.69}{683}100 = 0.25\%.$$

As we see, the black body heated to a temperature of 2,000 K is not efficient as a light source for the visible range. In reality, however, as mentioned in Chapter 3 (Section 3.1), thermal sources have a color temperature of about 3,000 K, which of course improves the luminous efficiency, but it is still very low.

10.3. The actual dynamic range, DR, can be defined as the ratio between the maximum illumination level, E_{max} (when saturation of the camera is achieved) and the minimum illumination level, E_{min} , indicated in the problem (0.3 lx). Evidently both illumination levels should be expressed in the same units (watts).

Let the area of a single pixel of the CCD be A' and the corresponding area of the illuminated scene be A, both areas being related as $A' = A(S'/S)^2$, where S is the distance from the camera lens to the object and S' is the distance from the lens to the CCD sensor. The latter can be replaced by the focal length of the lens, f', because usually $S \gg S'$. Then the energy coming from area A to a single pixel of the CCD is expressed as follows:

$$E_{px,\min} = E_{\min}A\frac{\omega}{2\pi} = E_{\min}A'\left(\frac{S}{S'}\right)^2\frac{\pi D^2}{4S^2}\frac{1}{2\pi} = E_{\min}A'\frac{1}{8(f/\#)^2}.$$
 (A)

To express E_{\min} in radiometric units (watts) we calculate the luminous efficacy, K, defined by Eq. (10.11) where the source radiant power in the integrals is related to sun radiation (sun color temperature T = 6,000 K) and the integral in the denominator covers the spectral range of the CCD sensitivity (0.4–1.1 µm).

We then have

$$K = 683 \frac{\int\limits_{0.4}^{0.77} e_{\rm B}(\lambda, T) V(\lambda) \,\mathrm{d}\lambda}{\int\limits_{0.4}^{1.15} e_{\rm B}(\lambda, T) \,\mathrm{d}\lambda} = 683 \frac{\Delta \lambda \sum_{i=1}^{8} a_i V_i}{\Delta \lambda \sum_{i=1}^{16} a_i}$$

where a_i are the spectral values from the second column of Appendix 3, V_i are the corresponding values of the eye spectral sensitivity from Fig. 10.6, and $\Delta \lambda = 0.05 \,\mu\text{m}$ for both integrals. The calculation yields $K = 132.98 \,\text{lm/W}$. With this value we get finally

$$DR = \frac{E_{\text{max}}}{E_{\text{min}}} = \frac{8.4 \times 10^{-2} \times 1.2^{2} \times 8 \times 132.98}{0.3} = 429$$

10.4. From the definition of complementary colors, we have, using Eq. (10.7)

 $x_{\rm c} = 0.33 - 0.2 = 0.13; \ y_{\rm c} = 0.33 - 0.16 = 0.17; \ z_{\rm c} = 1 - x_{\rm c} - y_{\rm c} = 0.7.$

This gives for the R, G, B components of an object with complementary color the following relations: $r/b = x_c/z_c = 13:7$; $g/b = y_c/z_c = 17:7$.

10.5. Calculation of *H*, *S*, *I* coordinates should be done according to Eq. (10.8). Before doing this it is useful to calculate the ratio of the R, G, B components of the object: r/b = x/(1-x-y) = 0.2/(1-0.2-0.5) = 2/3; g/b = y/(1-x-y) = 0.5/(1-0.2-0.5) = 5/3. By substituting these values in Eq. (10.8) we obtain:

$$H = 120^{\circ} + \arctan\left[\frac{\sqrt{3}(1 - g/b)}{2(r/b) - 1 - (g/b)}\right] = 120^{\circ} + \arctan\left[\frac{\sqrt{3}(1 - 5/3)}{4/3 - 1 - 5/3}\right]$$
$$= 160.89^{\circ} = 0.894\pi$$
$$S = \vartheta = \arctan\frac{\sqrt{3(1 - 5/3)^2 + (4/3 - 1 - 5/3)^2}}{\sqrt{2}(2/3 + 1 + 5/3)} = \arctan\frac{\sqrt{68}}{10}$$
$$= 39.5^{\circ} = 0.219\pi.$$

Therefore, the hue and saturation of the object are defined uniquely. As to intensity, it obviously depends on illumination level:

$$I = b \frac{r/b + g/b + 1}{\sqrt{3}} = b \frac{10}{3\sqrt{3}} = 1.925b.$$

10.6. The standard C phosphors are characterized by color coordinates x, y, z mentioned in Section 10.2. Therefore, each phosphor emits not a pure color but a

mixture of red, green, and blue. Let the brightness of the red phosphor be denoted as $B_{\rm R}^{(r)}$, $B_{\rm R}^{(g)}$, $B_{\rm R}^{(b)}$ for these three colors. The symbols for the brightness of the green and blue phosphors are similar. Furthermore, if the signals on the monitor input are $i_{\rm R}$, $i_{\rm G}$, and $i_{\rm B}$ for the three color channels the tristimulus values of radiation energized by these signals can be expressed as follows: $X_{\rm R} = i_{\rm R} B_{\rm R}^{(r)}$; $Y_{\rm R} = i_{\rm R} B_{\rm R}^{(g)}$; $Z_{\rm R} = i_{\rm R} B_{\rm R}^{(b)}$ and therefore

$$x_{\rm R} = \frac{X_{\rm R}}{X_{\rm R} + Y_{\rm R} + B_{\rm R}} = \frac{i_{\rm R} B_{\rm R}^{(\rm r)}}{i_{\rm R} B_{\rm R}^{(\rm r)} + i_{\rm R} B_{\rm R}^{(\rm g)} + i_{\rm R} B_{\rm R}^{(\rm b)}} = \frac{B_{\rm R}^{(\rm r)}}{B_{\rm R}^{(\rm r)} + B_{\rm R}^{(\rm g)} + B_{\rm R}^{(\rm b)}};$$
$$y_{\rm R} = \frac{B_{\rm R}^{(\rm g)}}{B_{\rm R}^{(\rm r)} + B_{\rm R}^{(\rm g)} + B_{\rm R}^{(\rm b)}}.$$

A similar relation can be evidently written for the two other phosphors. When all three phosphors of the screen are activated simultaneously the overall brightness is just the sum of that of the separate channels, and this is true for red and green and blue. Hence, the normalized color coordinates of the screen, x_S and y_S , are expressed in the following manner:

$$x_{\rm S} = \frac{B_{\rm R}}{B_{\rm R} + B_{\rm G} + B_{\rm B}} = \frac{B_{\rm R}^{(\rm r)} + B_{\rm G}^{(\rm r)} + B_{\rm B}^{(\rm r)}}{B_{\rm R}^{(\rm r)} + B_{\rm G}^{(\rm r)} + B_{\rm B}^{(\rm r)} + B_{\rm R}^{(\rm g)} + B_{\rm G}^{(\rm g)} + B_{\rm G}^{(\rm$$

By substituting the problem data in these two formulas we find the color coordinates of the screen as follows:

$$x_{\rm s} = \frac{(0.635 \times 0.8) + (0.305 \times 0.3) + 0.155}{2.1} = 0.359;$$

$$y_{\rm S} = \frac{(0.34 \times 0.8) + (0.595 \times 0.3) + 0.07}{2.1} = 0.248.$$

10.7. Since the imaging of the color scene on the CCD sensor is aimed at direct color measurement, with no involvement of any observation of the display, gamma-correction of the CCD signals is not required and we should choose $\gamma = 1$ for the camera electronics. The signals at the camera output will be determined from Eq. (10.13) where integration over the spectrum is reduced to the single values related to the monochromatic wavelength of 0.595 µm. From the graphs of Fig. 10.8 for the camera primaries we find $R_D(0.595) = 1.665$; $G_D(0.595) = 0.5$; $B_D = 0$ and therefore $i_R = 1.665A_R$; $i_G = 0.5A_G$; $i_B = 0$ where A_R and A_G are the

overall amplification factors of the camera electronics in red and green channels. These amplification factors are dictated by the camera calibration for white balance, i.e., the camera signals are equalized when the scene is illuminated by the light source of 3,200 K temperature:

$$i_{\rm WB}^{\rm (R)} = A_{\rm R} \int_{0.4}^{0.7} e_{\rm B}(\lambda, T) R_{\rm D}(\lambda) \, \mathrm{d}\lambda = i_{\rm WB}^{\rm (G)} = A_{\rm G} \int_{0.4}^{0.7} e_{\rm B}(\lambda, T) G_{\rm D}(\lambda) \, \mathrm{d}\lambda$$

To find the amplification factors one should calculate the integrals in the above equation using the black body radiation function from Appendix 3 and the camera primaries from Fig. 10.8. Numerical integration performed with 15 spectral intervals of 0.02 μ m each yields $I_{\rm R}$: $I_{\rm D} = 85.98 : 57.37 = A_{\rm G}/A_{\rm R}$. Hence

$$\frac{i_{\rm R}}{i_{\rm G}} = \frac{1.665A_{\rm R}}{0.5A_{\rm G}} = 1.592; \quad x = \frac{i_{\rm R}}{i_{\rm R} + i_{\rm G} + i_{\rm B}} = \frac{1.592}{1.592 + 1} = 0.614;$$
$$y = \frac{i_{\rm G}}{i_{\rm R} + i_{\rm G} + i_{\rm B}} = \frac{1}{2.592} = 0.386.$$

The last two values represent the measured color coordinates of the scene. In order to find the *H*-coordinate of the scene one should use the first expression of Eq. (10.8) which can be rewritten here as follows:

$$H = 120^{\circ} + \arctan\left(-\frac{\sqrt{3}i_{\rm G}}{2i_{\rm R} - i_{\rm G}}\right) = 120^{\circ} - \arctan\left[\frac{\sqrt{3}}{2(i_{\rm R}/i_{\rm G}) - 1}\right]$$
$$= 120^{\circ} - \arctan\frac{\sqrt{3}}{2 \times 1.592 - 1} = 81.6^{\circ} = 0.453\pi.$$

- W. Bachalo, C.F. Hess, and C.A. Hartwell, Trans. ASME 102, 798 (1980).
- M. Born and E. Wolf, Principles of Optics, Pergamon Press, 1968.
- R.M. Boynton, Human Color Vision, Holt, Rinehart and Winston, 1979.
- S. Brown (editor), *Mechanical Signature Analysis: Theory and Applications*, Academic Press, 1986, Chapter 12.
- G. Buchsbaum, Proc. IEEE 69, 772 (1981).
- F. Durst, Trans. ASME 104, 284 (1982).
- R.E. Hopkins, Chapter 4, in *Optical Design. Military Standardization Handbook*. Defense Supply Agency, Washington, 1962.
- A.F. Inglis, Video Engineering McGraw-Hill, 1993.
- R.S. Keyes (editor), *Topics in Applied Physics, Vol. 19: Optical and Infrared Detectors*, Springer-Verlag, 1977.
- R. Kingslake, Fundamentals of Lens Design, Academic Press, 1979.
- Oriel Instruments Catalogue, The Book of Photonics Tools, SpectraPhysics, 2003.
- V.C. Smith and J. Pokorny, Vision Research 12, 2059 (1972).
- W.J. Smith, Modern Optical Engineering, McGraw-Hill, 1984.
- G. Wyszecki and W.S. Stiles, *Color Science: Concepts and Methods, Quantification Data and Formulae*, John Wiley, 1982, 2nd edition.
- A. Yariv, An Introduction to Theory and Applications of Quantum Electronics, John Wiley, 1982.
- A. Yariv, Optical Waves in Crystals, John Wiley, 1984.
- M. Young, Optics and Lasers, Springer Series in Optical Science, 1984, Vol. 5.

This page intentionally left blank

Appendices

Appendix 1. Physical Constants

Constant	Symbol	Value	Units	
Planck's constant	h	6.6262×10^{-34}	J s	
Boltzmann's constant	k	1.3806×10^{-23}	$ m J~K^{-1}$	
Stefan–Boltzmann constant	σ	5.6696×10^{-8}	${ m W}{ m m}^{-2}{ m K}^{-4}$	
Speed of light in vacuum	с	2.999×10^{8}	${ m m~s^{-1}}$	
Electron charge	е	1.602×10^{-19}	С	
Electron mass	me	9.110×10^{-31}	kg	
Energy of 1 electron volt	eV	1.602×10^{-19}	J	
Energy of a photon at wavelength 0.5 μm	$E_{\rm ph}$	3.973×10^{-19}	J	
Avogadro's number	Ν	6.0222×10^{23}	mol^{-1}	
Volume of 1 gram-molecule	V_{μ}	22.42	1	
Universal gas constant	R	8.3170	$J K^{-1} mol^{-1}$	
Glass type	n _D	<i>n</i> _F	n _C	v
------------	----------------	-----------------------	----------------	--------
BK7	1.5168	1.52238	1.51432	64.12
K5	1.52249	1.52910	1.51982	56.30
F1	1.62588	1.63932	1.62074	33.686
SF5	1.67270	1.68876	1.66661	30.37
SF11	1.78472	1.80645	1.77599	25.76
SF57	1.84666	1.87425	1.83651	22.434

Appendix 2. Selected Data for Schott Optical Glasses

		177) <i>T</i>
λT	$\frac{e_{\rm B}(\lambda,T)}{\sigma T^5} 10^5$	$\int_{0}^{\lambda T} e_{\rm B}(\lambda, T) \mathrm{d}\lambda$	λT	$\frac{e_{\rm B}(\lambda,T)}{\sigma^{T_5}}10^5$	$\int_{0}^{\lambda T} e_{\rm B}(\lambda, T) \mathrm{d}\lambda$
μm.K	$(\mu m.K)^{-1}$	σT^4	μm.K	$(\mu m.K)^{-1}$	σT^4
800	0.03117	0.00001	5,500	10.342	0.6909
900	0.12767	0.00009	5,600	9.940	0.7011
1,000	0.3727	0.0003	5,700	9.553	0.7108
1,100	0.8559	0.0009	5,800	9.183	0.7202
1,200	1.6474	0.0021	5,900	8.827	0.7292
1,300	2.7757	0.0043	6,000	8.486	0.7378
1,400	4.223	0.0078	6,100	8.159	0.7462
1,500	5.9353	0.0129	6,200	7.845	0.1542
1,600	7.8294	0.0198	6,300	7.544	0.7618
1,700	9.8149	0.0286	6,400	7.257	0.7692
1,800	11.804	0.0394	6,500	6.981	0.7764
1,900	13.721	0.0524	6,600	6.717	0.7832
2,000	15.504	0.0668	6,700	6.464	0.7898
2,100	17.122	0.0831	6,800	6.220	0.7961
2,200	18.532	0.1009	6,900	5.987	0.8022
2,300	19.727	0.1202	7,000	5.766	0.8081
2,400	20.707	0.1404	7,100	5.553	0.8138
2,500	21.468	0.1615	7,200	5.349	0.8193
2,600	22.038	0.1832	7,300	5.153	0.8245
2,700	22.419	0.2055	7,400	4.966	0.8295
2,800	22.63	0.2280	7,500	4.787	0.8344
2,900	22.696	0.2506	7,600	4.614	0.8391
3,000	22.634	0.2733	7,700	4.451	0.8437
3,100	22.457	0.2959	7,800	4.291	0.8480
3,200	22.184	0.3182	7,900	4.141	0.8522
3,300	21.832	0.3402	8,000	3.995	0.8564
3,400	21.413	0.3618	8,500	3.354	0.8747
3,500	20.947	0.3830	9,000	2.832	0.8901
3,600	20.433	0.4036	9,500	2.404	0.9032
3,700	19.893	0.4237	10,000	2.0522	0.9143
3,800	19.325	0.4434	10,500	1.7606	0.9238
3,900	18.75	0.4624	11,000	1.5182	0.9320
4,000	18.161	0.4809	11,500	1.3153	0.9391
4,100	17.571	0.4987	12,000	1.145	0.9452
4,200	16.977	0.5160	12,500	1.000	0.9506
4,300	16.39	0.5327	13,000	0.878	0.9553

Appendix 3. Black Body Radiation

continued

	$e_{\rm P}(\lambda T)$	λT $\int e_{\mathbf{P}}(\lambda T) d\lambda$		$e_{\rm P}(\lambda T)$	$\lambda T \int \rho_{\rm PD}(\lambda T) d\lambda$
λT	$\frac{c_{\rm B}(\pi,T)}{\sigma T^5} 10^5$	<u>0</u>	λT	$\frac{\sigma B(n, 1)}{\sigma T^5} 10^5$	<u>0</u>
μm.K	$(\mu m.K)^{-1}$	σT^4	μm.K	$(\mu m.K)^{-1}$	σT^4
4,400	15.814	0.5488	13,500	0.773	0.9594
4,500	15.239	0.5643	14,000	0.684	0.9630
4,600	14.683	0.5793	14,500	0.607	0.9662
4,700	14.138	0.5937	15,000	0.540	0.9691
4,800	13.607	0.6075	20,000	0.196	0.9857
4,900	13.093	0.6209	25,000	0.0869	0.9923
5,000	12.594	0.6337	30,000	0.0441	0.9954
5,100	12.11	0.6461	35,000	0.0247	0.9911
5,200	11.643	0.6580	40,000	0.0149	0.9981
5,300	11.192	0.6694	45,000	0.00949	0.9986
5,400	10.758	0.6804	50,000	0.00634	0.9990

Appendix 3. continued

Material	Temperature (°K)	Emissivity, ε (normal, total)	
Metals			
Aluminum	1,000	0.054	
Copper	1,000	0.018	
Platinum	1,000	0.107	
Gold	1,000	0.025	
Nickel	1,000	0.128	
Iron	700	0.28-0.5	
Tungsten	1,300	0.131	
	3,300	0.4–0.8	
Refractories			
Brick	300	0.93	
Alumina	300	0.5	
Asbestos	400	0.96	
Concrete	1,000	0.63	
Dielectrics			
Glass	300	0.92	
Fused silica	300	0.93	
Water	300-400	0.95	

Appendix 4. Emissivity of Selected Materials

This page intentionally left blank

Index

A

Abbe invariant, 6 number, 45 Aberration, 5 Astigmatism, 51-52, 82 chromatic, 44, 75, 280 distrortion, 54 field curvature, 53 lateral, 41 offense against sine condition (OSC), 57, 84 plot, 43 rules of addition, 59 spherical, 49, 280 of cylinder lens, 51 third order, 49 transverse, 41 wave, 43 Absorptance, 210, 223 Absorption factor, 168, 208 Accuracy, sub-pixel, 148 Achromat (doublet lens), 45, 76 Acousto-optical cell (AOM), 233, 273, 275 deflector, 237, 238 effect, 233 modulator, 237 for spectral analysis, 239-241, 249-250 Airy's function, 62, 175 Anamorphic prism pair, 110, 125

Aperture angle, 19 stop, 19 Aplanatic points, 57, 85–87 Approach paraxial, 5

B

Ball lens, 10, 32, 58, 85 Bandpass, 187 Beam, optical, 2 convergent, 2 divergent, 2 Gaussian, propagation, 104 homocentric, 2 Beam expansion, 105 Beer's law, 193 Bending parameter (of a lens), 46 Bernoulli distribution, 133 Black body, 210, 214 Blazing, 180 angle, 180 Bolometer, 141 Bouguer's law, 167 Bragg cell, 234 condition, 235 Brightness, 212

С

Charge Coupled Device (CCD), 144–147 Color, complementary, 297, 306 coordinates, 296, 297, 301 detectors, 300 mixture curves, 295 perception, 294 x-y diagram, 296 Computed Radiography (CR), 170–171 Condenser, 60 Cone cells, 293 Cosine error, 256 Cross talk, 256 Cylinder lens, ray tracing, 78–82

D

Dark current, 131 Dark field illumination, 113, 126 Defocusing, 74 Detectivity, 130 specific (D-star), 131 Detectors, array, 143 CCD. 144-147 CMOS. 147 four quadrant, 143 semiconductor, photoconductive, 140 photovoltaic (photodiode), 140 thermal, 141 two-element, 143 Diffraction, 61, 174 angle, 174 limiting system, 61 grating, plane, 178 reflective, 180 optimization, 183 concave, 184 Directional ambiguity, 273 Dispersion, angular, 169 linear, 169 Display screens, 303 phosphors, 303, 306-307 primaries, 303 Doppler broadening of spectral lines, 161 Dual path arrangement (for AOM), 236, 246 Dynamic range, 132 Dynodes, 137

Ε

Emissivity, 210 Emittance, 210 Encircled energy distribution, 67 Error, location, 242 Eye, human, 10–11 angular resolution, 11, 32 spectral sensitivity, 248

F

F-center, 166 f-number, f#, 51 Fabry-Perot interferometer, 188, 206 Field lens, 16 Field of view, 20 Field stop, 20 Filter, interference, 186, 205 multi-cavity, 188, 206 IR cut-off. 147. 300 Flattener lens, 56, 84 Flow, laminar, 274 turbulent, 269 two-phase, 281 Focal length, 5 back. 5 front, 5 measurement, 9, 28 Focus, 5 Fourier spectrometer, 190-193, 207 Fresnel's formulas, 167 Frequency, cut-off, 69 hopping, 111 shifting, 273, 285 Fringe pattern, 270, 277 spacing, 270, 285 Full width at half maximum (FWHM), 187, 205.206

G

Galvanometer scanner, 231–232 Gamma-correction procedure, 303 Gear profile measurement, 258–259 Gladston–Dale formulae, 233 Geometrical optics assumptions, 3 signs convention, 3

Η

Homocentricity violation, 10, 30

I

Illumination system, lens-based, 97 single lens, 98 two-lens, 98, 116 three-lens, 99, 118 oblique, 113, 127 of a microscope (opaque illuminator), 95, 117 Image formation, 4 Image, ideal, 5 Imaging, 1 by graphical method, 8, 25–26 Inclination angle, 180 Intensity (of radiation), 212 Interference, multi-beam, 178 Interferometer, dual beam, 218 Fabry-Perot, 188, 206 laser, 254–255 Irradiance, 95

Κ

Kirchhoff's law, 210

L

Lambert's law, 212 Lamp, arc, xenon, 96 Deuterium, 96 Quartz Tungsten Halogen (QTH), 95-96 Laser beam, divergence angle, 102 Gaussian profile, 102 waist, 102 Laser diode, 109 system for ground profiling, 111, 123-125 Laser Doppler velocimetry (LDV), 269 signal burst, 271, 272, 281-282 Laser-guided robot, 108 Laser light, 101 modes, 106, 110-111 Laser reference system for construction, 109 LDV arrangement, Forward scattering, 272 Back scattering, 273, 276 Side scattering, 274, 278 Light Emitting Diode (LED), 112 Line generator, 115 Line scanner, 238 Luminescence, 164-166 photostimulated (PSL), 166, 170 Luminous efficacy, 209, 304, 306 Luminous efficiency, 299, 304

Μ

Magnification, optical, 6 angular, 11 linear, 11 longitudinal, 12 visible, 12 Magnifier, simple, 12–13 Microscope, 13 magnification, 14 diffraction theory of imaging, 64 trinocular, 15 with ICS optics, 15 Modulation, 68 Modulation Transfer Function (MTF), 68–69, 92–93 Monochromator, 168, 172

Ν

Nernst rod, 95, 97 Noise, read-out, 135 shot, 133 thermal (Johnson), 134 Noise Equivalent Power (NEP), 130

0

Optical constants, 167, 170 Optical path difference (OPD), 43–44, 180, 183, 218 Optically active material, 101 Optical resonator, 101, 106 Optical tracking, 147, 152

Р

Paralax error, 71 Particle sizing, 281 Petzval's theorem, 54 Photoelectric cell, 136 Photomultiplier, 137 Photopigment, 293 Plane, principal, 7, 9, 29 Plank's law, 211 p-n junction, 140 Point Spread Function (PSF), 68 Poisson distribution, 133 Polygon, 230-231, 233, 245-246 Primaries, natural, 294 of the detectors, 300, 307 Prisms, 22 Amici (roof prism), 23 dispersive, 173 Dove, 23 minimum deviation angle, 38 penta, 23

Prisms (continued) rhomboidal, 37 right angle, 23 unfolded diagram, 24, 36 Probe volume, 270, 277, 285, 286 Pupil, Entrance/Exit, 19 Pyrometer, 214, 217

Q

Quantum efficiency, 130 of luminescence, 166, 170

R

Rangefinder, 251, 259 Reflectance, 167, 210 Reflection law, 2-3 Refraction law, 3 Refractive index, spectral behavior, 45 of air, 252 Ray, optical, 2 Ray tracing, 3, 7, 26-27 for cylinder lens, 51 Raster, 231 Rayleigh's criteria, wave aberration, 44 limiting resolution, 62 spectral resolution, 169 length, 120 Relay lens, 13-14 Resolution, 62 spectral, 169 Responsivity, 130 Reticle, 87 Retroreflector, 255 Reynolds' stresses, 269 Rod cells, 293 Rodopsin, 293 Rowland circle, 184, 205

S

Saturation, 132 Scanner, mirror, 229, 232, 242 fast rotating, 230, 232, 245, 253, 261–264 galvanometric, 231 acousto-optical, 237 Scanning error, exposure, 243 location, 242, 245 misalignment Scattering angular diagram, 283 Seeing, 293 Seidel's formula, 49 Signal-to-noise Ratio (SNR), 130 Slit, entrance, 172 optimal width, 173 Solar radiation, 163 Spatial filter, 106 Spectral analysis of electrical signals (RF), 239-241 Spectral imaging, 236, 247-248 Spectral resolution, 173 Spectrum, absorption, 162 of fused silica, 168 emission, 159 luminescence, 164 secondary, 47, 77 tertiary, 47 Spectrometer, 168 autocolimating, 183 Fabry-Perot, 188 Fourier, 190-193 prism-based, 172 with diffractive grating, plane, 182 concave, 184 Spectrophotometer, 168, 194 Stefan-Boltzmann law, 212 Stoke's rule, 165 Stratified light, 97, 257, 267-268 Surveying system, 104

T

Telecentric system, 71, 93 Telephoto lens, 72, 94 Telescope, 15-16 Galilean, 17 Temperature, brightness, 216, 223 Color, 96, 215, 223 radiation. 215-216 Temperature gradients measurement, 218 Thermal radiation, 209 laws, 210-211 sources, 95, 305 Three-CCD color sensor, 302 Time response, 131 Time-Bandwidth product (TBW), 238, 249 Transmittance, 167 Tristimulus generalization, 294-295

Index

U

Units, photometric, 298

V

Video signal, standard, 145 Vignetting, 21 Visibility function, 282

W

Wave number, 160 White balance, 303, 308 Width, natural, of spectral line, 160 Wien's law, 211 Wien's formula, 213, 222–223 Wobbling error, 230 Work function, 136