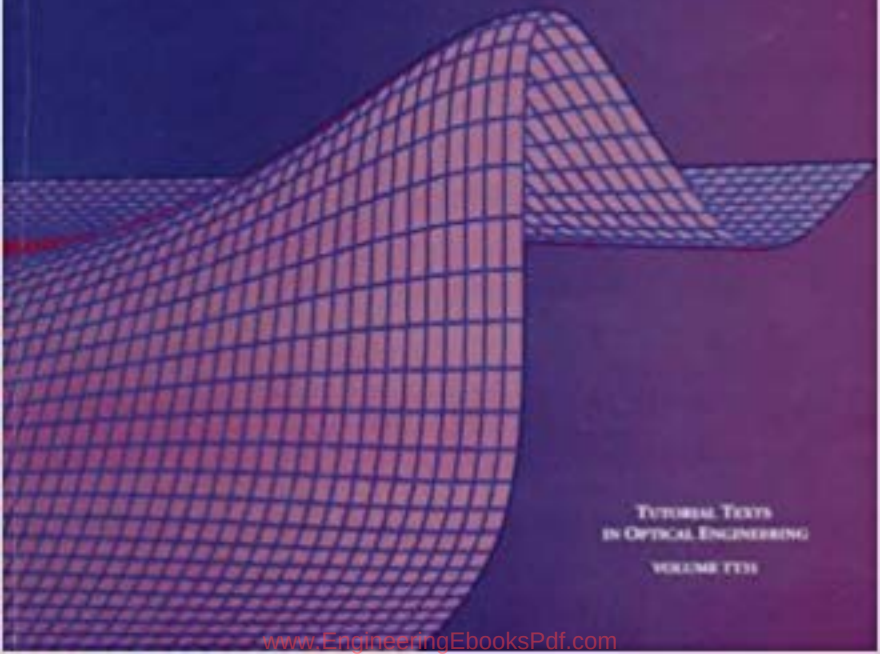
Basic Electro-Optics for Electrical Engineers

GLENN D. BOREMAN



TUTORIAL TEXTS
IN OPTICAL ENGINEERING
VOLUME TT31

www.EngineeringEbooksPdf.com

Basic Electro-Optics for Electrical Engineers

Glenn D. Boreman

Tutorial Texts in Optical Engineering

Volume TT31

SPIE Optical Engineering Press

A Publication of SPIE—The International Society for Optical Engineering

Bellingham, Washington USA

www.EngineeringEbooksPdf.com

Library of Congress Cataloging-in-Publication Data

Boreman, G. D. (Glenn D.)
Basic electro-optics for electrical engineers / Glenn D. Boreman.
p. cm. – (Tutorial texts in optical engineering; v. TT31)
Includes bibliographical references and index.
ISBN 0-8194-2806-X (pbk.)
1. Integrated optics. 2. Electrooptical devices.
3. Optoelectronic devices. I. Title. II. Series.
TA1660.B67 1998
621.36—dc21 97-52690
CIP

Published by

SPIE—The International Society for Optical Engineering
P.O. Box 10
Bellingham, Washington 98227-0010
Phone: 360/676-3290
Fax: 360/647-1445
Email: spie@spie.org
WWW: http://www.spie.org/

Copyright © 1998 The Society of Photo-Optical Instrumentation Engineers

All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means without written permission of the publisher.

Printed in the United States of America.

www.EngineeringEbooksPdf.com

Copyrighted Materials
Copyright © 1998 SPIE Retrieved from www.knovel.com

Preface

This book is an outgrowth of a short course frequently given over the past ten years for The International Society for Optical Engineering (SPIE). Basic optical-engineering concepts are covered in a manner accessible to anyone with a preparation equivalent to a bachelor's degree in electrical engineering. Students with backgrounds in mechanical engineering, technology, and management have also taken the course successfully.

The topics covered are those most likely to be useful in an entry-level, laboratory-oriented setting. These include: imaging, radiometry, sources, detectors, and lasers. Special emphasis is placed on flux-transfer issues, which are particularly important in practical measurements. A first-order approach is taken throughout, so that the student can quickly make the back-of-the-envelope calculations necessary for initial setup of optical apparatus. The treatment is at an elementary conceptual and computational level.

The material contained herein represents some of what I wished I had known about optics when I first began working as an engineer, fresh from The University. Over the intervening years, I have developed the opinion that, while valuable for advanced students, such topics as the eikonal equation or the method of stationary phase are not as important for the young engineer as answers to questions like: where is the image, how big is it, how much light gets to the detectors, and how small of an object can we see?

I would like to thank the students who have taken my SPIE short courses on this material. Their questions and enthusiasm have provided the impetus for a number of revisions of the presentation. Special thanks are also due to Eric Pepper of SPIE, who has shown remarkable patience with innumerable schedule slippages. His unfailing optimism and cheerfulness are sincerely appreciated.

Finally, I want to thank my favorite technical editor, Maggie Boreman, for the time that she has invested in this project, transforming my convoluted prose into standard English.

GDB
Geneva, Florida, 1998

www.EngineeringEbooksPdf.com

Copyrighted Materials
Copyright © 1998 SPIE Retrieved from www.knovel.com

Table of Contents

www.EngineeringEbooksPdf.com

www.EngineeringEbooksPdf.com

Chapter 1
Geometrical Optics

1.1 Introduction

Geometrical optics is the study of image formation. The function of lens and mirror systems is to collect radiation from the object and produce a geometrically similar distribution of flux, an image, across the image plane.

We consider first the electromagnetic spectrum, then proceed to the principles of image formation. We then consider issues affecting image quality and conclude this chapter by surveying the optical properties of materials.

1.2 Electromagnetic Spectrum

Light is a sinusoidal electromagnetic wave that travels either in a material medium or in a vacuum. We describe electromagnetic (EM) radiation using wavelength (λ), defined as the distance over which a sinusoidal waveform repeats itself. We will consider many optical phenomena that depend on wavelength. Radiation classified according to wavelength falls into one of the categories listed in Table 1.1. The divisions between the various categories are approximate and based on differences in source and detector technologies, the response of the human eye, and atmospheric transmission. Optical radiation is only a small portion of the entire EM spectrum, which extends from radio waves at long λ to x rays at short λ. Electromagnetic radiation to which our eyes respond has wavelengths between approximately 0.4 μm (violet) and 0.7 μm (red). Invisible radiation includes ultraviolet (UV) radiation, which has wavelengths shorter than violet, and infrared (IR) radiation, which has wavelengths longer than red.

Table 1.1. Electromagnetic spectrum.

Wavelength (λ)	Designation
10^{-6} to 10^{-2} μm	x rays
10^{-2} to 0.4 μm	ultraviolet
0.4 to 0.7 μm	visible
0.40 - 0.46	violet
0.46 - 0.49	blue
0.49 - 0.55	green
0.55 - 0.58	yellow
0.58 - 0.60	orange
0.60 - 0.70	red
0.7 μm to 3 μm	near infrared
3 μm to 5 μm	middle infrared
5 μm to 15 μm	long-wave infrared
15 μm to 1 mm	far infrared
1 mm to 1 cm	millimeter waves
1 cm to 1 m	microwaves
1 m to 10 km	radiowaves

www.EngineeringEbooksPdf.com

The wavelength is related to the oscillation's optical frequency (ν) according to

$$\nu = c/\lambda \tag{1.1}$$

where $c = 3 \times 10^8$ m/s is the speed of light in vacuum (or air). As the wave travels throught various media, the wavelength will change. However, the frequency of the radiation will not change, once the wave has been generated. Thus, a decrease in velocity requires a corresponding decrease in wavelength. The index of refraction (n) is the proportionality constant, and is defined as the ratio of the wave's vacuum velocity to the wave velocity in the material (v)

$$n = c/v \tag{1.2}$$

Because an EM wave has its maximum velocity in vacuum, the refractive index is a number greater than one for any material medium.

Another relationship involving wavelength and frequency is

$$\mathcal{E} = hc/\lambda = h\nu \tag{1.3}$$

where $\mathcal{E}$ is the photon energy, and $h = 6.67 \times 10^{-34}$ J s is Planck's constant.

For infrared radiation at $\lambda = 10$ μm, Eq. (1.3) predicts a frequency of $\approx 3 \times 10^{13}$ Hz and a photon energy of $\approx 2 \times 10^{-20}$ J. At a shorter wavelength of 1 μm, the frequency and the photon energy both increase by a factor of ten, to yield $\nu = 3 \times 10^{14}$ Hz and $\mathcal{E} = 2 \times 10^{-19}$ J. A typical visible wavelength of 0.6 μm has a frequency on the order of 5×10^{14} Hz. Because present-day electronics does not permit direct temporal measurements on time scales shorter than approximately 10^{-9} seconds, at optical and IR frequencies we cannot measure directly the amplitude and phase of the EM wave. Measurements can be made only on average power, averaged over many cycles of the wave.

1.3 Rays

A fundamental concept in geometrical optics is that of a ray. A ray is a line along which energy flows from a source point. The ray model is an approximation that ignores the effects of diffraction, and thus assumes that light travels in straight lines. The ray model is a useful starting point for analysis, because of its simplicity. We can calculate image size and image location using the ray model, and later include the effect of diffraction in the design process. Because we want to see how objects and images can be constructed on a point-by-point basis, we will begin our discussion of rays with a point source of radiation.

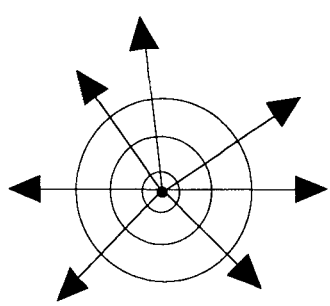


Figure 1.1. Rays from a point source.

www.EngineeringEbooksPdf.com

A point source emits rays uniformly in all radial directions. In Fig. 1.1, wavefronts can also be seen. The wavefronts represent families of points that are at a given time of flight (phase value) from the source. The wavefronts for a point source are concentric circles centered on the source. Note that the rays and the wavefronts are everywhere perpendicular.

Optical systems are often configured to form an image of a distant target. In drawing the ray diagram for such a system, we need a construction for a point source at infinity. Modifying the construction seen in Fig. 1.1, we find in Fig. 1.2 that the wavefronts eventually become planar, at least sufficiently planar over a limited region of space. Any optical system will necessarily have a finite aperture, so the very small curvature of the wavefront from the source can be neglected. As the wavefronts become planar, the associated rays become parallel. Very often, we see optical systems drawn with entering rays that are parallel. These rays come from a point source which is "at infinity."

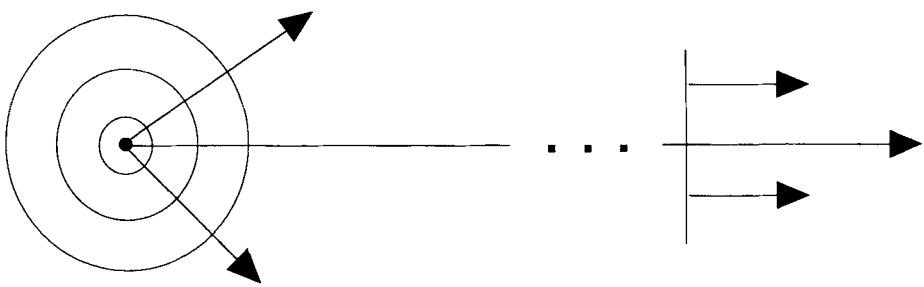


Figure 1.2. A point source at infinity is a source of parallel rays.

1.4 Imaging Concepts

We will concentrate on what is called first-order (paraxial) optics, an approximation which assumes that angles made by rays with the horizontal axis are sufficiently small that the sine of an angle can be replaced by the angle itself in radians. This linearizes the raytrace equations that determine ray paths through the optical system, and allows us to predict image size and image location. A more detailed analysis is necessary for prediction of image quality, involving the tracing of nonparaxial rays at larger angles.

Referring to Fig. 1.3, an object is a spatial distribution of brightness. The optic axis is a line of symmetry for the lens that contains the centers of curvature of the lens surfaces. Each location in the object can be considered an independent source point that emits light rays into all forward directions. The object information is that each source point can have a different strength. The lens causes a geometrically similar distribution of brightness to be reproduced in the image. The action of the lens ensures that rays diverging from a particular object point will intersect again at a corresponding point in the image plane, building the image point by point. The brightness at any image location is proportional to the brightness of the corresponding object point, so that the image is a distribution of flux that is geometrically similar to that of the object. Given that the number of object points is infinite (each with an infinite number of rays that could be drawn), we want to simplify our diagrams, drawing only a few rays from selected object points.

www.EngineeringEbooksPdf.com

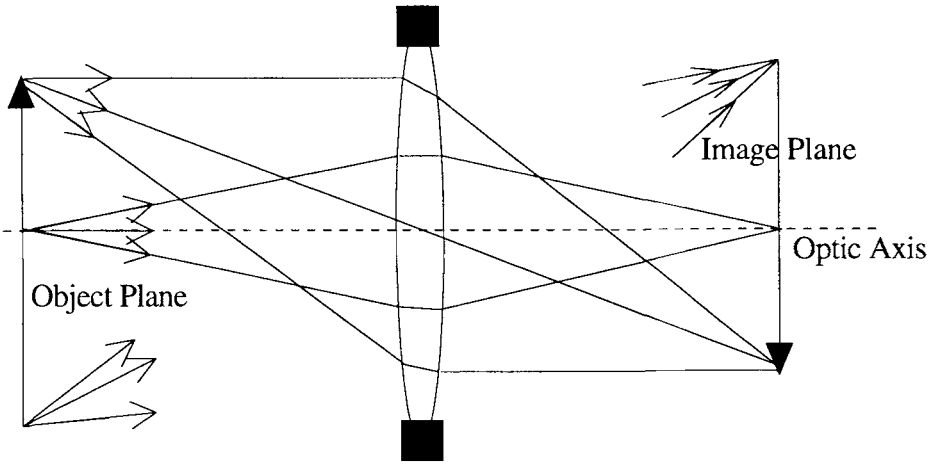


Figure 1.3. Image formation.

To obtain the maximum insight from just a few rays, we choose the rays that we want to draw using the graphical raytrace rules shown in Fig. 1.4. In this figure, we assume that the lens is thin. The thin-lens approximation assumes that the lens thickness along the optical axis is negligible compared to its focal length, object distance, or image distance; or that any ray has only a very small change in height while traversing the lens. The graphical raytrace rules applicable to thin lenses are:

1. Rays entering the lens parallel to the optic axis exit through a focal point.
2. Rays entering the lens through a focal point exit parallel to the optic axis.
3. Rays that pass through the center of the lens do not change direction.

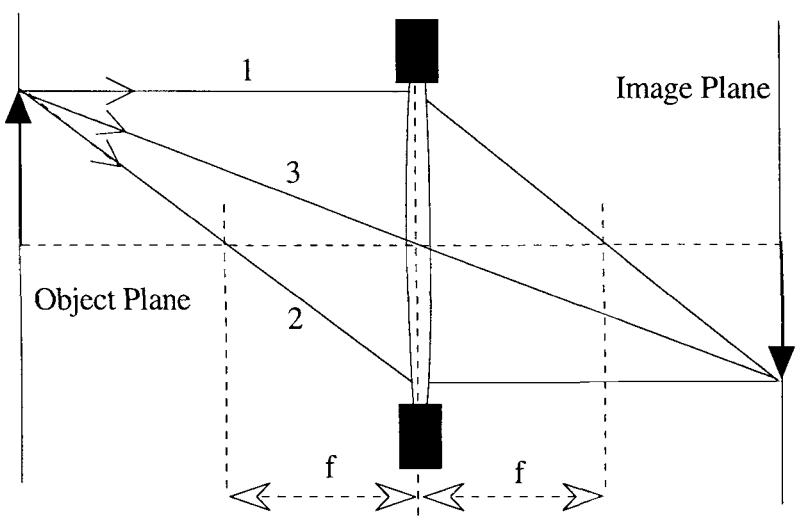


Figure 1.4. Graphical raytrace rules.

A lens has a focal point on each side, as seen in Fig. 1.5, and parallel rays entering from the left will converge to the focal point on the right side of the lens.

www.EngineeringEbooksPdf.com

Similarly, rays that diverge from the focal point on the left side of the lens will emerge parallel to the optic axis. The construction in Fig. 1.5 justifies Rules 1 and 2. For a thin lens, we measure the focal length from the lens itself to either of the focal points. The focal length is the same on either side, provided that the propagation medium (usually air) is the same on either side. Focal length is positive for the lens in Fig. 1.5.

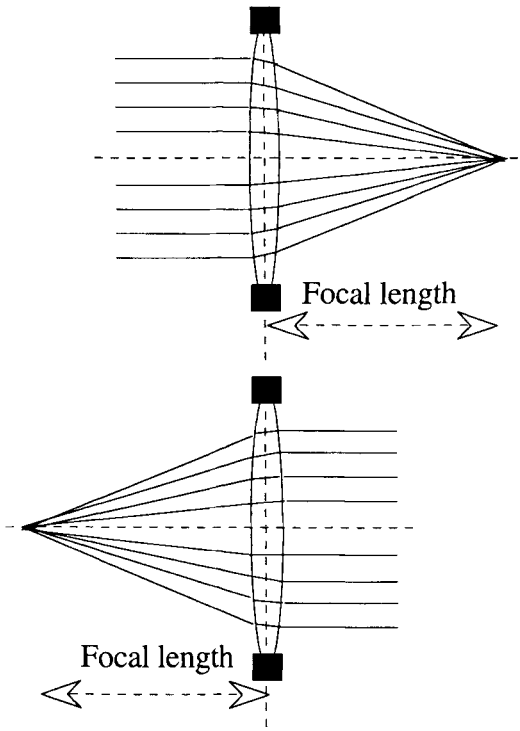


Figure 1.5. Definition of focal length for a thin lens.

Rule 3 can be justified considering the region of the lens near the axis as approximately a plane parallel plate as seen in Fig. 1.6. By Snell's Law, the ray passing through this region will be shifted parallel to its original direction. For a thin lens, the shift is small, and the ray appears to go straight through the center of the lens.

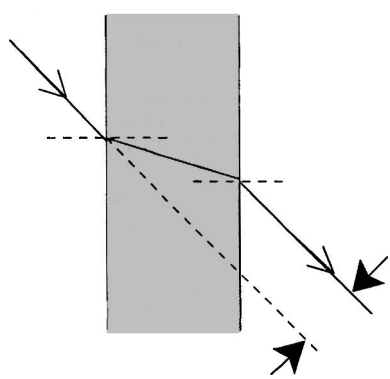


Figure 1.6. Raytrace Rule 3.

www.EngineeringEbooksPdf.com

A useful application of the graphical raytrace rules is seen in Fig.1.7, in which a point source at infinity creates a parallel bundle of rays that make an angle α (in radians) with respect to the optic axis. Assuming that α is small, the bundle of rays will focus at a distance αf above the axis, and at a distance f to the right of the lens.

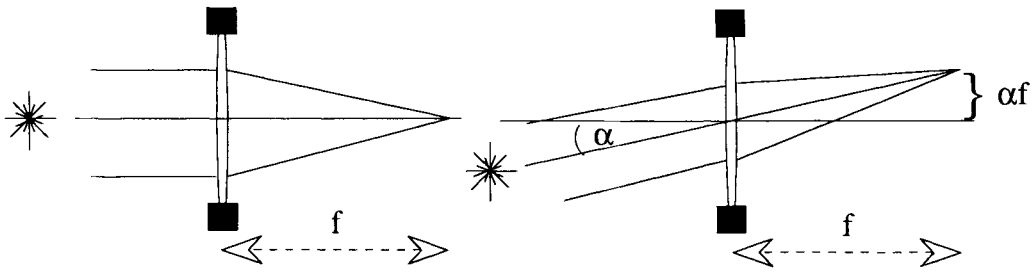


Figure 1.7. Off-axis ray bundle.

The graphical raytrace rules can be used to develop a pair of algebraic relations for thin-lens imaging:

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f} \tag{1.4}$$

and

$$\mathcal{M} = \frac{h_{img}}{h_{obj}} = -\frac{q}{p} \quad . \tag{1.5}$$

As seen in Fig. 1.8, the distance p is that from the object to the lens (p is defined to be positive to the left of the lens) and the distance q is that from the lens to the image (q is defined to be positive to the right of the lens). The ratio of the image height (h_{img}) to the object height (h_{obj}) defines the linear magnification $\mathcal{M}$. Both heights are defined to be positive above the optic axis, resulting in a minus sign in Eq. (1.5) corresponding to the inverted image formed by a single positive lens.

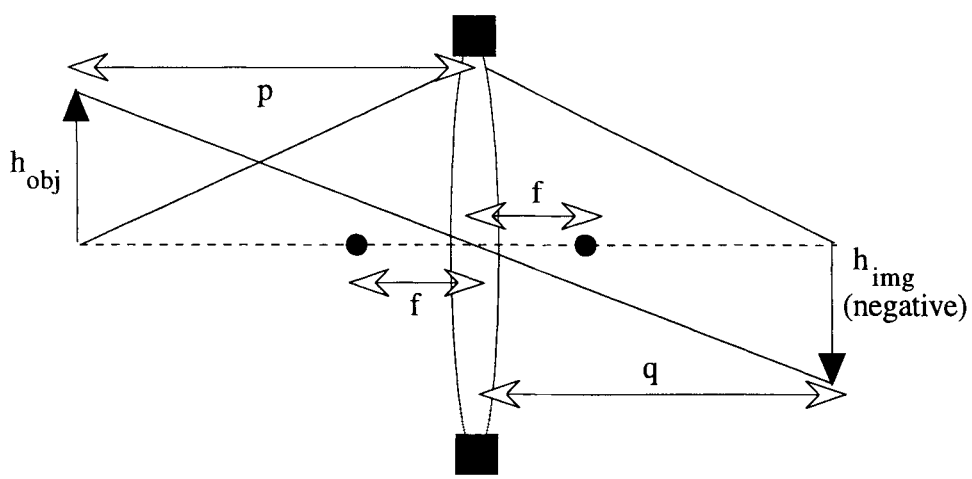


Figure 1.8. Imaging equation for single thin lens; all quantities except h_{img} are positive.

www.EngineeringEbooksPdf.com

Using Eqs. (1.4) and (1.5), we can calculate the (paraxial) image location and image size, given the object location, object size, and the focal length of the lens. As an example, given a lens with a focal length of 5 cm, and a 3-mm-tall object located 20 cm to the left of the lens, we find q = 6.67 cm and $\mathcal{M} = -0.33$, yielding an image that is 1 mm in dimension below the axis. Usually objects and images are centered on the optic axis, but using just the dimension above or below the axis allows us to keep track of the image orientation as well.

Let us examine some limiting cases of Eqs. (1.4) and (1.5). If the object is located at infinity on the left, then the image is formed at the focal point of the lens on the right. As the object comes in from infinity, the image is formed at an increasing distance to the right of the lens. As the object is moved toward the left-hand-side focal point, the image will be formed at infinity on the right-hand side.

We can also use Eq. (1.4) to show that the distance $\ell \equiv p + q$ between real object and real image (p and q both positive) has a minimum value of $\ell = 4f$. For the symmetrical case where p = q = 2f, Eq. (1.5) yields $\mathcal{M} = -1$. For an object-to-image distance $\ell > 4f$, we use the quadratic equation to find two pairs of p and q that satisfy Eq. (1.4) and produce reciprocal magnifications:

$$p = \frac{\ell}{2} \pm \frac{\sqrt{\ell^2 - 4\ell f}}{2} \quad . \tag{1.6}$$

In the previous situation, we had p = 6.67 cm, and q = 20 cm, for a total length $\ell = 26.67$ cm and a magnification $\mathcal{M} = -0.33$. Another solution with the same total length is p = 6.67 and q = 20, which yields $\mathcal{M} = -3$. The magnification $\mathcal{M}$ defined in Eq. (1.5) is a linear magnification, and if the areas of the object and image are the quantities of interest, the area magnification $\mathcal{M}^2$ is used, as seen in Fig. 1.9

$$\frac{A_{img}}{A_{obj}} = \frac{(2\,h_{img,y})(2\,h_{img,x})}{(2\,h_{obj,y})(2\,h_{obj,x})} = \mathcal{M}^2 = \left(\frac{q}{p}\right)^2 \quad . \tag{1.7}$$

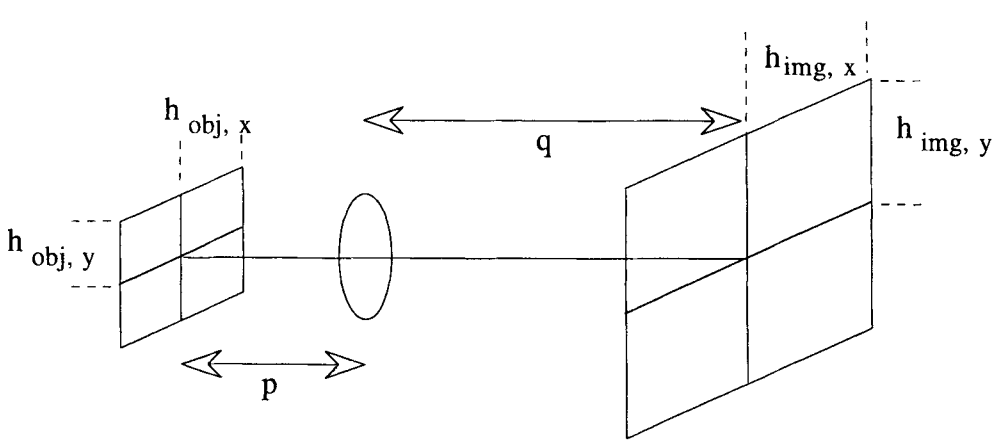


Figure 1.9. Area magnification.

www.EngineeringEbooksPdf.com

1.5 Aperture Stop, Marginal Ray, and F/#

The size and location of apertures in an optical system affect both the flux-transfer efficiency and the image quality of the system. We begin by considering an axial ray, seen in Fig. 1.10. An axial ray starts at the on-axis point in the object and ends at the on-axis point in the image. The aperture stop is the surface that limits the acceptance angle for axial rays. The marginal ray is that particular axial ray that goes through the edge of the aperture stop.

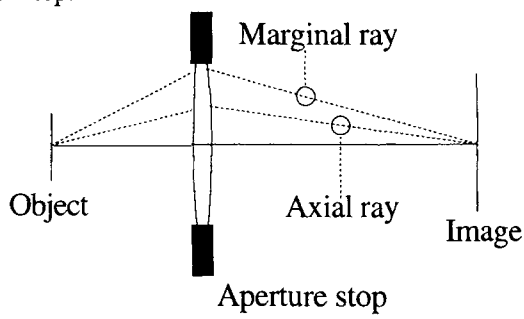


Figure 1.10. Axial ray, aperture stop, and marginal ray.

The marginal ray is used to define the relative aperture or f-number (F/#), which compares the aperture diameter to the object distance, image distance, or focal length. The F/# can be defined in one of three ways, depending on the situation under consideration. If the object is at infinity (Fig. 1.11), the image is formed at the focus of the lens, according to Eq. (1.4). In this case, the marginal ray defines the image-space F/# as

$$(F/\#)_{\text{image-space}} = f/D_{\text{aperture-stop}} \cdot \tag{1.8}$$

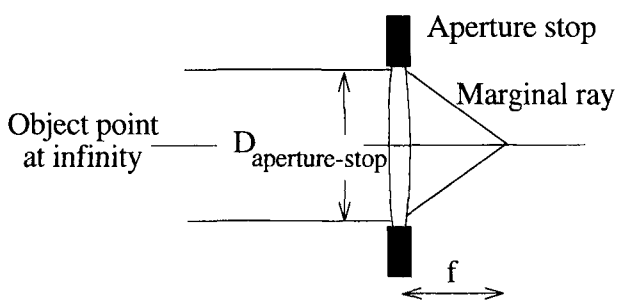


Figure 1.11. F/# for object at infinity.

For objects that are at finite conjugates (object not at infinity) as seen in Fig. 1.12, an F/# can be defined either in object or image space as

$$(F/\#)_{\text{object-space}} = p/D_{\text{aperture-stop}} \tag{1.9}$$

$$(F/\#)_{\text{image-space}} = q/D_{\text{aperture-stop}} \cdot \tag{1.10}$$

www.EngineeringEbooksPdf.com

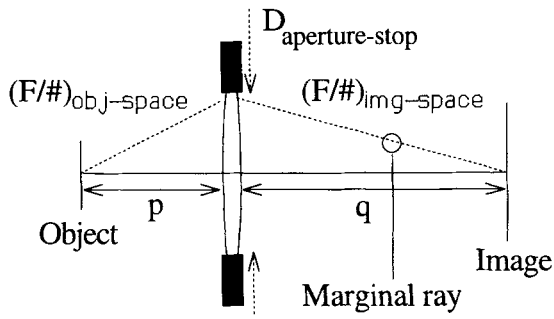


Figure 1.12. F/# for object at finite conjugates.

F/# is a parameter that depends on the inverse of the aperture-stop diameter. Hence, a large F/# implies that less relative aperture is used. A system with a large F/# is termed "slow" and one with a small F/# is termed "fast," referring to the relative exposure speeds that can be used with each system. We will see later that a fast system produces a higher image brightness than a slow system. We compare F/1 (F/# = 1) and F/3 (F/# = 3) systems in Fig. 1.13, where F/1 implies that $f = D_{aperture\text{-}stop}$ and F/3 implies that $D_{aperture\text{-}stop} = (1/3) \times f$.

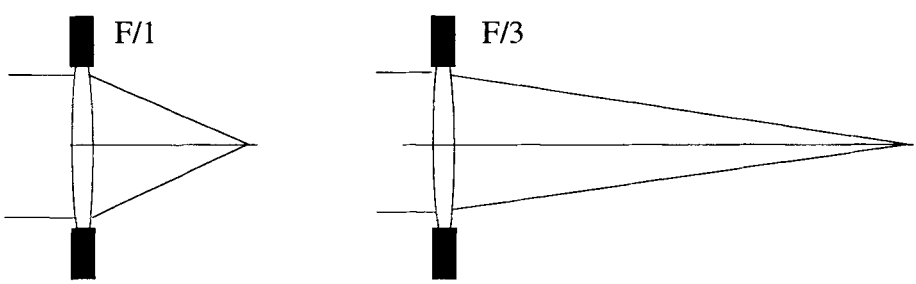


Figure 1.13. A comparison of F/1 and F/3 systems.

1.6 Field Stop, Chief Ray, and Field of View

Another important ray in an optical system is the chief ray. As seen in Fig. 1.14, the chief ray starts at the maximum extent of the object and goes to the maximum extent of the image, and passes through the center of the aperture stop. The maximum extent of the object and image is determined by a surface called the field stop. Without a field stop, there would be an infinite extent of the object plane that would be imaged onto the image plane. The portions of the image that were formed at a considerable distance from the optic axis would have poor image fidelity; they would be blurry. In any real imaging system, a field stop is necessary to define the region of image space over which an image will be recorded with a given degree of image quality. The field stop is usually determined by the limited size of the recording medium (e.g., photographic film, a detector array, or a single detector) placed in the image plane.

www.EngineeringEbooksPdf.com

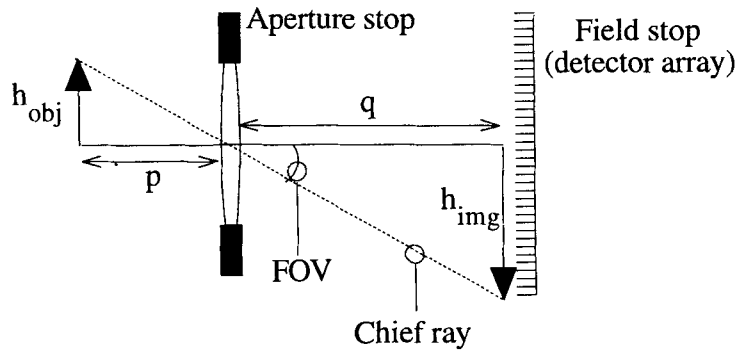


Figure 1.14. Field stop, chief ray, and field of view.

The field of view (FOV) is the angular coverage of an optical system, and can be defined in terms of either a full angle or a half angle. Using the half-angle FOV convention of Fig. 1.14, we find

$$\text{FOV}_{\text{half-angle}} = \left|\tan^{-1}\left(\frac{h_{obj}}{p}\right)\right| = \left|\tan^{-1}\left(\frac{h_{img}}{q}\right)\right| . \tag{1.11}$$

The angles used to define FOV in Eq. (1.11) are equivalent because of graphical raytrace Rule #3, which says that the chief ray is not changed in angle because it passes through the center of the aperture stop.

We usually think of the object as being imaged onto the image plane. A useful alternative construction is available when the image plane acts as the field stop with a detector array located there, as seen in Fig. 1.14. We consider the image of the detector array in object space. Any individual detector will have a "footprint" in object space. The footprint is that region of object space that contributes flux to (is imaged onto) the detector of interest. Figure 1.15 illustrates a specific case that represents the typical situation in remote-sensing applications. When the object is sufficiently far away that the image is formed at the focus of the lens ($p \approx \infty$ implies that $q \approx f$), the linear dimension of the footprint in the object plane is

$$y_{\text{footprint}} = \frac{y_d\, p}{f} \tag{1.12}$$

where y_d is the linear dimension of the individual detector. Surely any linear dimensions in the object and image plane are related by Eq. (1.5), but Eq. (1.12) applies to the condition where the object is essentially at infinity.

For the object-at-infinity case it is also of interest to relate FOV and magnification. As seen in Fig. 1.16, the (full angle) FOV = y_{array}/f, and for a given detector array, a long-focal-length lens yields a narrow FOV while a short-focal-length lens yields a wide FOV. In this figure, the footprint of the detector array (for a given lens) will appear as full width on the video display. This relationship between FOV and magnification can be seen from the relation of Eq. (1.5), $\mathcal{M} = -q/p$, and, because $q \approx f$, a longer focal length implies a longer image distance which yields a higher magnification.

www.EngineeringEbooksPdf.com

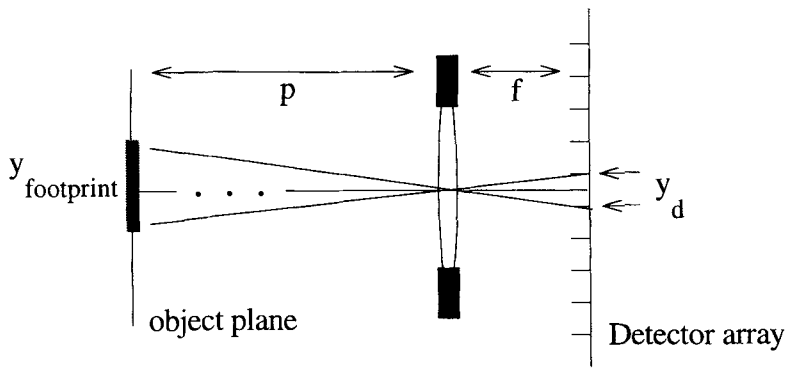


Figure 1.15. Detector footprint.

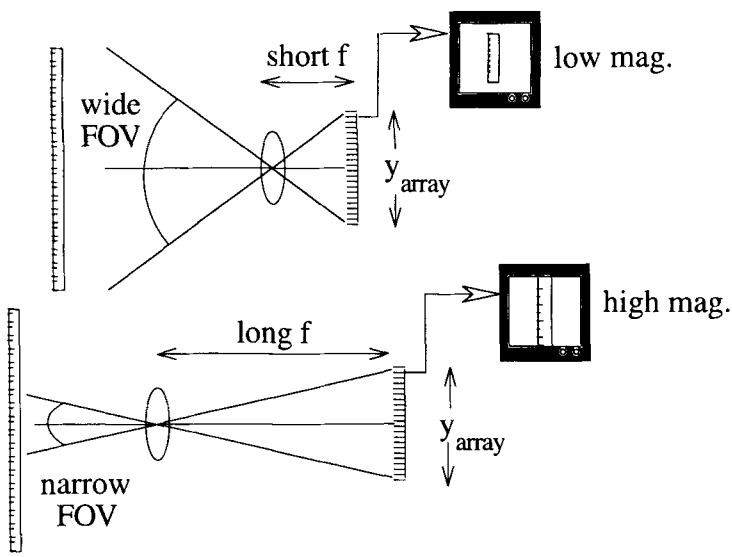


Figure 1.16. Field of view, magnification, and focal length.

In the last two sections, we have explored two special rays, the marginal ray and the chief ray. These rays define two parameters of the optical system, the F/# and FOV, respectively. The F/# and the FOV affect both flux transfer and image quality. A small F/# and large FOV both promote a high flux-transfer efficiency, but the image quality suffers (large aberrations). In contrast, a system with a large F/# and a small FOV will have better image quality (smaller abberations), but less flux reaching the image plane.

1.7 Thick Lenses and Lens Combinations

Equations (1.4) and (1.5) describe the paraxial imaging for thin lenses, that is, for lenses where the focal length can be described simply as the distance from the lens itself to the focal point. To describe imaging for optical elements with appreciable thickness (compared to their focal length), or for combinations of thin lenses, we develop

www.EngineeringEbooksPdf.com

the concept of principal planes. A pair of these planes acts as equivalent refracting planes for a given optical system. For thick lenses, the focal length must be measured from the focal point to the principal point (rather than to the lens itself). The procedure for locating the principal points of a thick lens is straightforward, and follows the steps illustrated in Fig. 1.17.

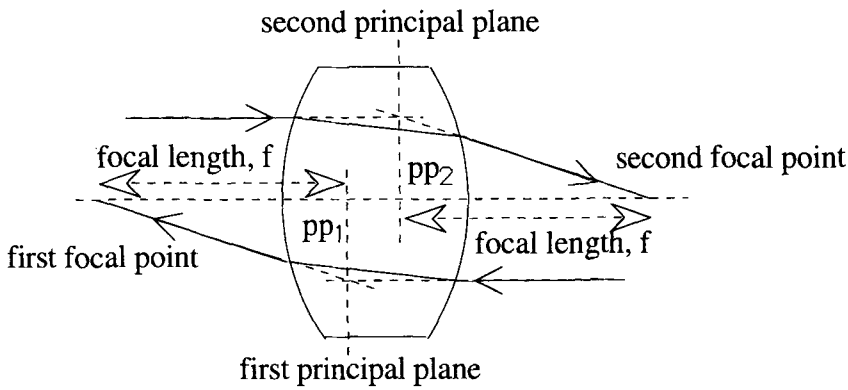


Figure 1.17. Principal planes.

The steps for location of the principal planes are:

1) Use a ray entering parallel to the optic axis from the right.
2) Project the entering ray and the exiting ray until they intersect.
3) Construct a perpendicular line to the optic axis.
4) Intersection of the perpendicular and optic axis is the first principal point pp_1.
5) Second principal point pp_2 can be found in a similar fashion, using a parallel ray entering from the left.

The focal length f is measured from the principal point to the corresponding focal point, and is the same on each side of the lens, provided that the index of refraction of object and image space is equal. The distance from the last physical surface of the lens to the focal point (called the back focal distance) is not necessarily the same on each side, but the focal length will be equal. For thin lenses, both principal points coincide at the lens.

Using the concept of principal planes, we can describe several useful combinations of thin lenses. The first combination we consider is two positive thin lenses, with focal lengths f_1 and f_2, separated by a distance d (Fig. 1.18). The focal length of the combination is:

$$f_{combination} = \left[\frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2} \right]^{-1} . \qquad (1.13)$$

As seen in Fig. 1.18, the focal length is measured from the appropriate principal plane, not from either of the individual lenses.

www.EngineeringEbooksPdf.com

As we let the separation approach zero, we obtain a formula for two lenses in contact:

$$\frac{1}{f_{combination}} = \frac{1}{f_1} + \frac{1}{f_2} \quad . \tag{1.14}$$

For example, two thin lenses in contact, each with a focal length of 3 cm, will result in a combination focal length of 1.5 cm.

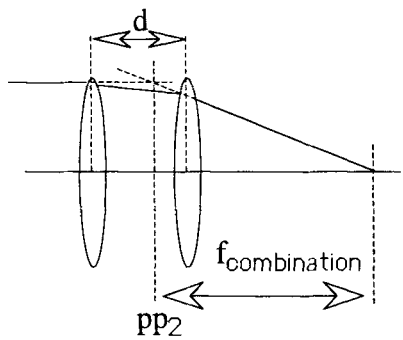


Figure 1.18. Two thin lenses, separated by distance d.

A useful special case of the two-lens combination is the relay-lens pair, seen in Fig. 1.19. The object is placed f_1 to the left of the first lens, and the image is formed at f_2 to the right of the second lens. The separation of the lenses affects both the location of the principal planes and the focal length of the combination. However, because the object is located at the front focal point of the lens #1, axial rays will be parallel to the optic axis in the space between the lenses. The lens separation will thus not affect the magnification of the relay, which is

$$\mathcal{M}_{relay} = \frac{h_{img}}{h_{obj}} = -\frac{f_2}{f_1} \quad . \tag{1.15}$$

For increasing interlens separation, the required diameter of the second lens will increase to allow for an object of finite size to be imaged without losing the edges (vignetting). Consider the chief ray from the top of the object to the bottom of the image in Fig. 1.19. As the lenses are moved apart, the size of lens #2 must increase to pass the chief ray.

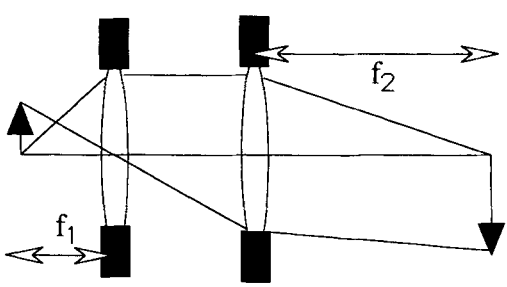


Figure 1.19. Relay-lens pair.

www.EngineeringEbooksPdf.com

Afocal telescopes are two-lens systems intended to be used with the object at infinity. The two lenses are separated by the sum of their focal lengths. The combination focal length is indeterminate using Eq. (1.13), which is interpreted to mean that an afocal telescope has no single-lens equivalent. A parallel marginal-ray bundle of diameter D_{input} is input from the left, and a parallel marginal-ray bundle of diameter D_{output} is output to the right. No single lens can accomplish this. Afocal telescopes do not produce an image by themselves, but provide an angular magnification of the marginal rays, $\mathcal{M}_{angular}$:

$$\mathcal{M}_{angular} \equiv \frac{\theta_{output}}{\theta_{input}} = -\frac{f_1}{f_2} \tag{1.16}$$

$$\left|\mathcal{M}_{angular}\right| = \frac{D_{input}}{D_{output}} = \frac{|f_1|}{|f_2|} \quad . \tag{1.17}$$

One basic afocal telescope is the Keplerian telescope seen in Fig. 1.20, which combines two lenses with positive focal length. The angular magnification for this telescope is negative, because both focal lengths are positive in Eq. (1.16).

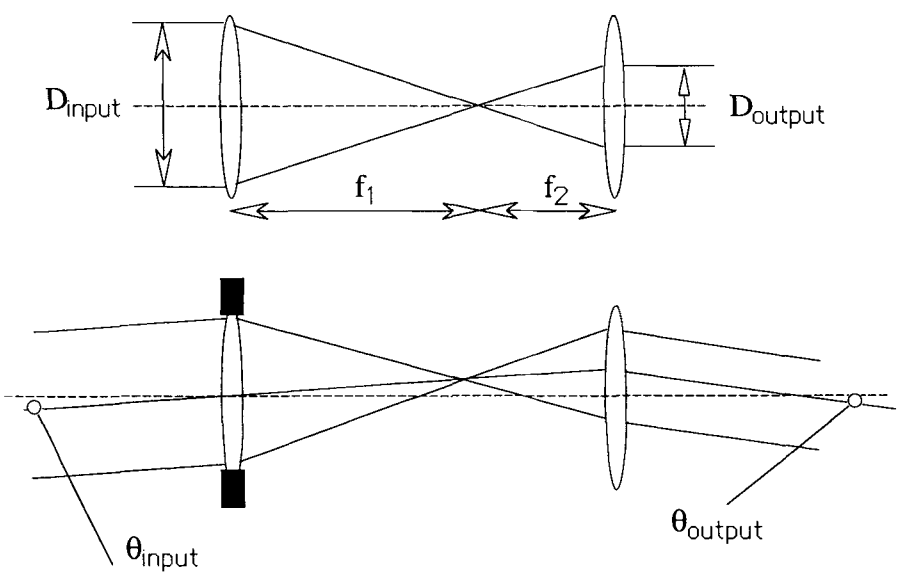


Figure 1.20. Keplerian telescope.

Another afocal-telescope configuration is the Galilean telescope, where the first lens is positive and the second lens is negative. The lens separation is the algebraic sum of the focal lengths. As seen in Fig. 1.21, using Eq. (1.16), the angular magnification for the Galilean telescope is positive. Galilean telescopes have a shorter physical length for a given angular magnification than the Keplerian configuration. A reflective analog to the Galilean telescope is seen in Fig. 1.22. The obstruction caused by the secondary mirror of this telescope makes the aperture stop an annulus. Slight adjustment of the spacing between the elements or the focal lengths yields a non-afocal configuration that can focus parallel rays called a Cassegrain system (Fig. 1.23).

www.EngineeringEbooksPdf.com

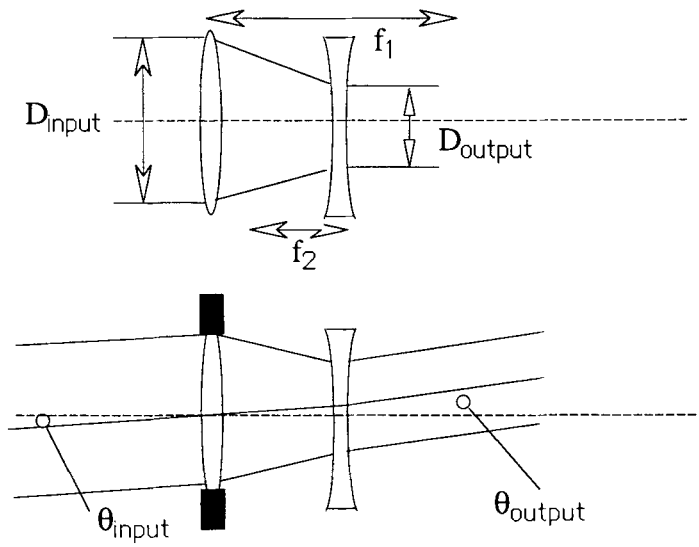


Figure 1.21. Galilean telescope.

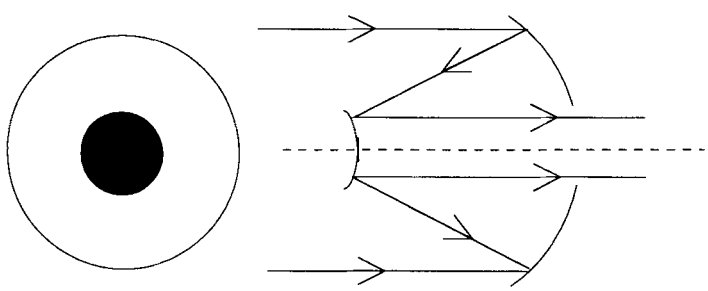


Figure 1.22. Reflective analog of the Galilean afocal telescope.

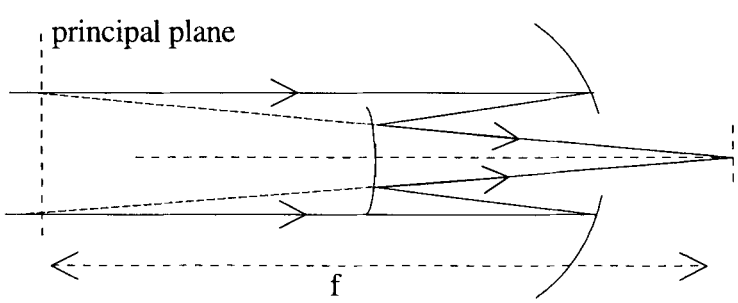


Figure 1.23. Cassegrain system.

www.EngineeringEbooksPdf.com

Afocal telescopes are typically used as the front-end element to modify the FOV of an image-forming system (Fig. 1.24). The afocal telescope does not form an image by itself, but acts in concert with a final positive lens (f_3) to form the final image on the detectors. The focal length of the three-lens combination is different from the original focal length of the imaging lens by a factor of the angular magnification (Eq. (1.16)) of the afocal. A change in the focal length of the system changes the FOV (and thus the magnification) of the system, as seen in Fig. 1.16. The focal length of the three-lens system (measured from the principal plane to the focal plane) is given by

$$f_{combination} = \frac{f_1 \times f_3}{f_2} = \left|\mathcal{M}_{angular}\right| \times f_3 \quad . \tag{1.18}$$

The system focal length can also be expressed as the focal length of the front lens of the afocal, multiplied by the magnification of the relay-lens pair formed by the second and third lenses

$$f_{combination} = \frac{f_1 \times f_3}{f_2} = f_1 \times \left|\mathcal{M}_{relay}\right| \quad . \tag{1.19}$$

The magnification of the afocal is chosen to give the final desired FOV, and can be conveniently changed for another afocal assembly on the front end of the system.

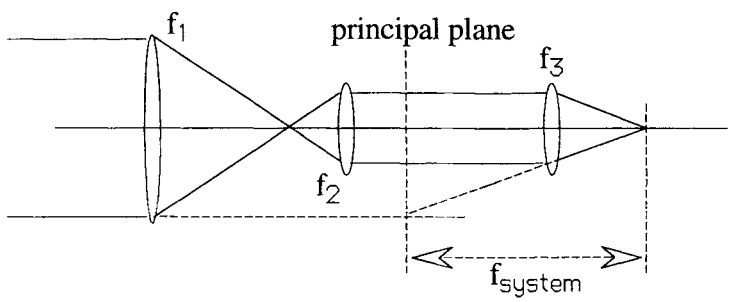


Figure 1.24. Three-lens combination of afocal telescope and imaging lens.

1.8 Image Quality and Diffraction

In our previous development, we made the paraxial assumption of perfect imaging: that is, points in the object map to points in the image. However, because of the effects of diffraction and aberrations, points in the object map to finite-sized blur spots in the image. Intuitively, better image quality is obtained for smaller blur spots.

Diffraction is caused by the wave nature of electromagnetic radiation. An optical system with a finite-sized aperture stop can never form a point image. Rather, a characteristic minimum blur diameter is formed, even when other image defects (aberrations) are absent. An imaging system limited only by the effects of diffraction is capable of forming a blur spot whose diameter is related to the image-space F/# of the marginal rays that converge to the ideal image point:

www.EngineeringEbooksPdf.com

$$d_{diffraction} = 2.4\,\lambda\,(F/\#)_{image\text{-}space} = \frac{2.4\,\lambda\,q}{D_{aperture\text{-}stop}} \tag{1.20}$$

$$d_{diffraction} = 2.4\,\lambda\,(F/\#)_{image\text{-}space} = \frac{2.4\,\lambda\,f}{D_{aperture\text{-}stop}} \quad \text{(for object at } \infty\text{)} \tag{1.21}$$

The actual blur-spot diameter formed by a particular system may well be larger because of aberrations. The diffraction-limited spot diameter quoted in Eqs. (1.20) and (1.21) is the best (smallest) spot size that an optical system can form, for a given wavelength and F/#. The diffraction pattern of a point source has some ring structure in it, as seen in Fig. 1.25, and the spot diameter quoted above is the diameter out to the first dark ring, which encloses 84% of the flux in the diffraction pattern. The functional form of the radial irradiance profile of the diffraction pattern for a sysem with a circular aperture is

$$E(r)\left[W/cm^2\right] = \left[\frac{2J_1\left(\frac{r}{\lambda\,(F/\#)}\right)}{\left(\frac{r}{\lambda\,(F/\#)}\right)}\right]^2, \tag{1.22}$$

proportional to the square of the Fourier transform of the aperture itself. The first dark ring of E(r) occurs at r = 1.2 λ F/#, where the J_1 Bessel function has its first zero.

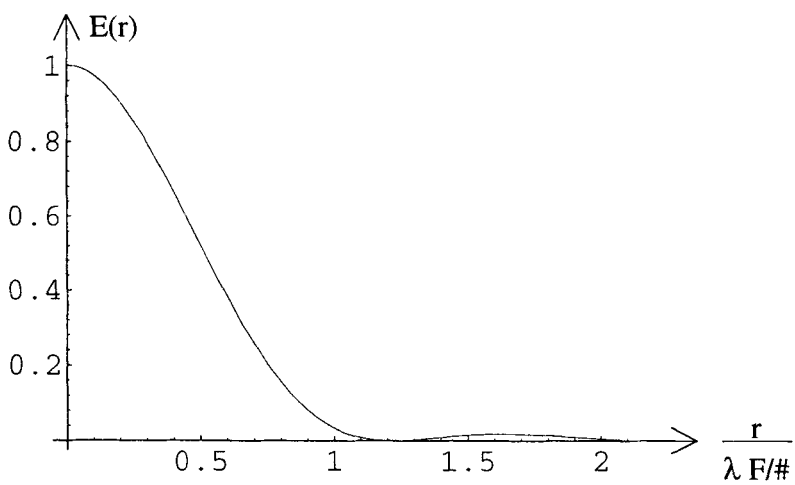


Figure 1.25. Diffraction pattern, radial irradiance profile: W/cm² vs. r/[λ(F/#)].

Let us consider a numerical example of a diffraction-limited F/2 system using infrared radiation of λ = 10 μm. This system is capable of forming a spot diameter no smaller than 48 μm. The same system operating at F/8 would form a spot diameter of 192 μm. It can be seen that low-F/# systems have the smallest diffraction-limited spot sizes, and thus the best potential image-quality performance. However, we will see in the next section that the effect of aberrations makes low-F/# systems more difficult to correct to a diffraction-limited level of performance.

www.EngineeringEbooksPdf.com

The effects of diffraction may also be expressed in angular terms, as seen in Fig. 1.26. For the case of a single thin lens with the stop at the lens, the diffraction-limited angular blur β depends only on the diameter of the aperture stop and the wavelength

$$\beta = \frac{2.4\,\lambda}{D_{\text{aperture-stop}}} \quad . \tag{1.23}$$

This angular blur can be multiplied by the image distance q (or f for an object at infinity) to yield a spot diameter consistent with Eqs. (1.20) and (1.21). The diffraction blur angle can also be multiplied by the object distance p to yield the footprint of the diffraction blur in the object plane. This footprint b× p represents the minimum resolution feature in the object.

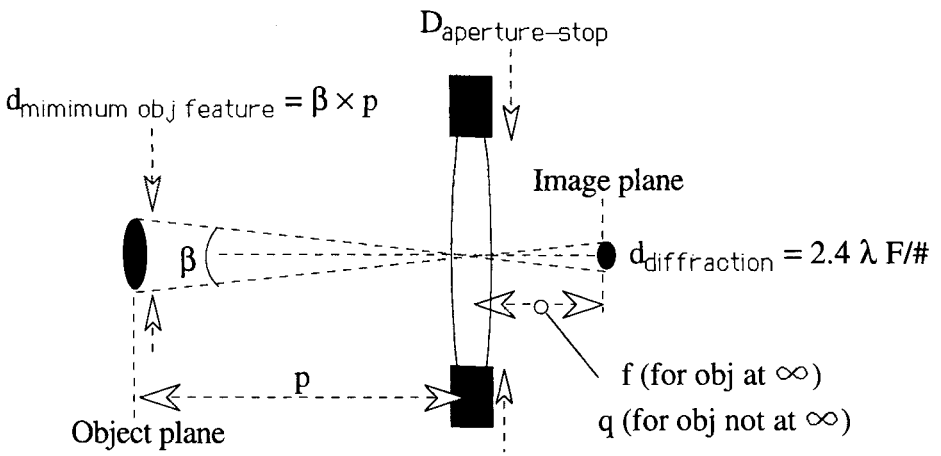


Figure 1.26. Diffraction-limited resolution angle.

1.9 Image Quality and Aberrations

Aberrations are image defects that arise from the deviation of rays from the ideal paraxial (small-angle) behavior. These deviations arise because an optical system is typically comprised of spherical surfaces. A system with a finite number of spherical surfaces cannot produce geometrically perfect imaging for all possible object points. The following aberrations are not fundamentally caused by defects in manufacture or alignment, but are inherent in systems having spherical surfaces: spherical aberration, coma, astigmatism, distortion, field curvature, and chromatic aberration. All of these image defects get worse with either decreasing F/#, increasing field angle, or both. A system that operates at a low F/# has the smallest diffraction-limited spot size, and thus the best potential performance. However, the effects of aberrations generally become worse (producing larger spot sizes) as the F/# decreases, so that a low-F/# system is harder to correct to a diffraction-limited level of performance.

Point sources in the object thus map to blur spots in the image, whose size depends on the F/# of the imaging lens and the field angle of the particular object point being considered. Control of aberrations in the design process generally dictates a minimum allowable F/# and a maximum allowable FOV. However, the design of any optical system must also consider the system's flux-transfer and field-coverage

www.EngineeringEbooksPdf.com

requirements. Thus, it can be seen that a fundamental tradeoff exists between the amount of flux in the image (which increases with increasing FOV and decreasing F/#) and the image quality.

Beyond restriction of field coverage and relative aperture, the optical designer achieves an optimization of image quality by a tradeoff process, exploiting two free variables in particular: the distribution of power between the surfaces of any element (lens bending) and the location of the aperture stop of the system. Neither of these variables affects the focal length of the lens, and both have large impact on image quality. Further optimization of image quality usually involves the creation of more degrees of freedom in the system (e.g., curvatures, thicknesses, indices of refraction). A more complex system, with smaller aberrations, is the result. In the optical-design process, correction of aberrations so that the aberration spot size is smaller than the spot size dictated by diffraction provides negligible benefit. The larger diffraction blur at infrared wavelengths makes it possible for infrared systems to be simpler than visible-wavelength imaging systems with similar aperture and field requirements.

1.10 Materials Considerations

The fundamental material property from an optics viewpoint is the refractive index (n), defined as the ratio between the velocity of an electromagnetic wave in free space to its velocity in the material under consideration. For most optical glasses, the refractive index ranges between 1.5 and 1.7. The refractive index of water is 1.33. Some infrared-transmitting materials have a high refractive index, such as silicon (n = 3.4) and germanium (n = 4.0).

An optical system will typically contain a number of interfaces between the various optical materials (including air) that comprise the system. These refractive-index boundaries cause some of the light to be reflected and some to be transmitted (refracted) at each interface. Both Snell's Law and the Fresnel equations depend on the difference in refractive index across a material boundary. The orientation of the refracted rays are determined by Snell's law, and the distribution of power between the reflected and refracted component is determined by the Fresnel equations. Hence, the refractive index of a material will affect both the direction and strength of the light rays that propagate into that material.

The refractive index of any material depends on the wavelength of the incident radiation. This property is called material dispersion. Figure 1.27 shows the characteristic s-shape of the refractive index-*vs*-wavelength curve.

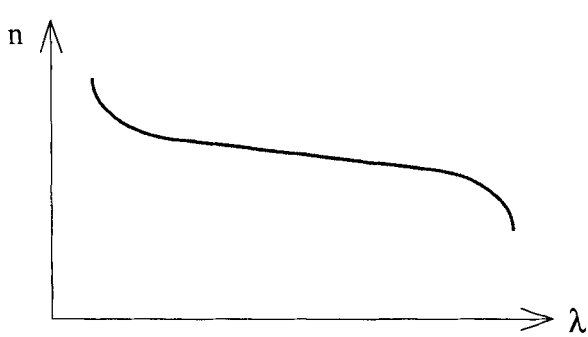


Figure 1.27. Material dispersion.

www.EngineeringEbooksPdf.com

Whenever a ray crosses a boundary between two materials having different refractive indices, the distribution of power between the transmitted and reflected components is determined by the Fresnel equations:

$$R = \left(\frac{n_2 - n_1}{n_2 + n_1}\right)^2 \tag{1.24}$$

$$T = \frac{4\, n_2\, n_1}{(n_2 + n_1)^2} \tag{1.25}$$

where n_1 and n_2 are the indices on each side of a boundary. The reflected power per surface (R), and the transmitted power per surface (T) are used as follows:

$$\phi_{reflected} = \phi_{incident} \times R \tag{1.26}$$

$$\phi_{transmitted} = \phi_{incident} \times T \tag{1.27}$$

where the symbol ϕ represents power (W). Equations (1.24) and (1.25) sum to unity, indicating that the absorption of the materials is assumed to be negligible. Another assumption inherent in Eqs. (1.24) and (1.25) is that they are approximations for small angle of incidence. At larger angles of incidence, R generally increases and T correspondingly decreases.

An example of the Fresnel equations can be found in Fig. 1.28, where the incident power is unity and the block of material is immersed in a medium of uniform refractive index. The reflected ray from the first surface has a strength R, with T being transmitted into the material. The ray reflected from the second surface is R times the ray strength in the material, or T × R. This leaves T × T transmitted back into the surrounding medium. If the block of material were immersed in air and had a refractive index typical of optical glass, R would be 4% and the overall transmitted power through the block would be 92%. However, if the block were made of a higher-index medium, say n = 3.0, then the power reflection coefficient R would be 25%, with an overall transmitted power of 56.25%. For such high-index materials (common in infrared applications), the use of thin-film antireflection coatings is necessary to reduce such large reflection losses.

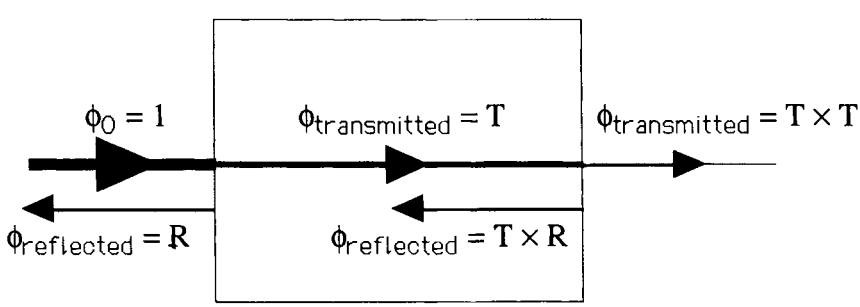


Figure 1.28. Fresnel equation example.

www.EngineeringEbooksPdf.com

Antireflection coatings operate on the principle of interference of waves, and are used to increase the transmittance of optical elements above that dictated by Eqs. (1.24) to (1.27). Referring to Fig. 1.29, a film of quarter-wavelength thickness

$$t_{film} = (\lambda/4n_{film}) \tag{1.28}$$

is deposited on the substrate material. Rays reflected from the top of the film will thus be 180° out of phase with the rays that have been reflected from the film-substrate interface, having travelled one round trip through the film. The phase shift between these two rays will allow them to interfere destructively to zero if their strengths are equal. If we approximate this condition (neglecting multiply-reflected rays inside of the film) by setting the power reflection coefficient R equal for the surround-film and film-substrate interfaces, we obtain the requirement

$$n_{film} = \sqrt{n_{surround}\, n_{substrate}} \ . \tag{1.29}$$

If the requirements of Eqs. (1.28) and (1.29) are satisfied, the two reflected rays will cancel, allowing all of the power in the incident ray to proceed into the substrate.

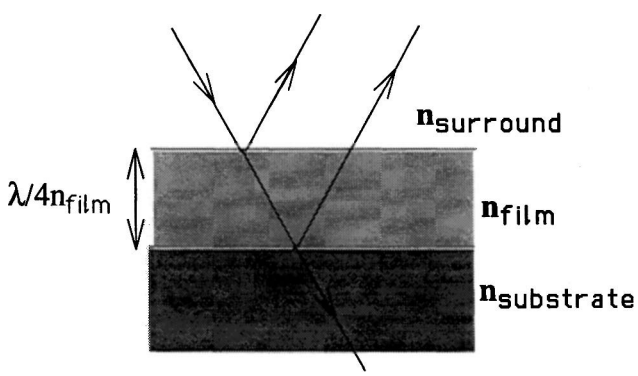


Figure 1.29. Single-layer antireflection coating.

Bibliography

Born, M., and E. Wolf, *Principles of Optics*, Pergamon, New York, 1975.
Hecht, E., and A. Zajac, *Optics*, Addison-Wesley, Reading, MA, 1974.
Jenkins, F., and H. White, *Fundamentals of Optics*, McGraw-Hill, New York, 1976.
Longhurst, R. S., *Geometrical and Physical Optics*, Longman, New York, 1973.
Smith, W. J., *Modern Optical Engineering*, McGraw-Hill, New York, 1966.

www.EngineeringEbooksPdf.com

Copyrighted Materials
Copyright © 1998 SPIE Retrieved from www.knovel.com

Chapter 2
Modulation Transfer Function

2.1 Introduction

An optical system's image quality can be characterized in the spatial domain using the impulse response (spot size) of the system, or in the spatial-frequency domain using the Fourier transform of the impulse response, the transfer function. An ideal system would form a point image of a point object. But, because of diffraction and aberrations, a real system has an impulse response of nonzero width. The impulse response of a system is the two-dimensional image formed in response to a delta-function object, and is denoted by the symbol h(x,y). The actual image, g, formed by the system is the ideal image, f (an exact replica of the object with appropriate size scaling), convolved with the impulse response, h:

$$f(x,y) ** h(x,y) = g(x,y) , \tag{2.1}$$

where the double asterisk denotes a two-dimensional convolution. A narrower impulse response gives better image quality than a wide impulse reponse.

Alternately, we can consider the imaging process in the spatial-frequency domain. In this context, we are concerned with the imaging of sinusoids of different frequencies, rather than the imaging of point objects. The irradiance distribution of an object can be thought of as composed of "spatial frequency" components, in the same way as a time-domain electrical signal is composed of various temporal frequencies by means of a Fourier analysis. First, consider an irradiance distribution as a function of x and y, as seen in Fig. 2.1. This irradiance distribution can represent either an object or an image. From a one-dimensional profile of the distribution, we obtain a single-variable function that can be Fourier decomposed into its constituent spatial frequencies (cycles/mm).

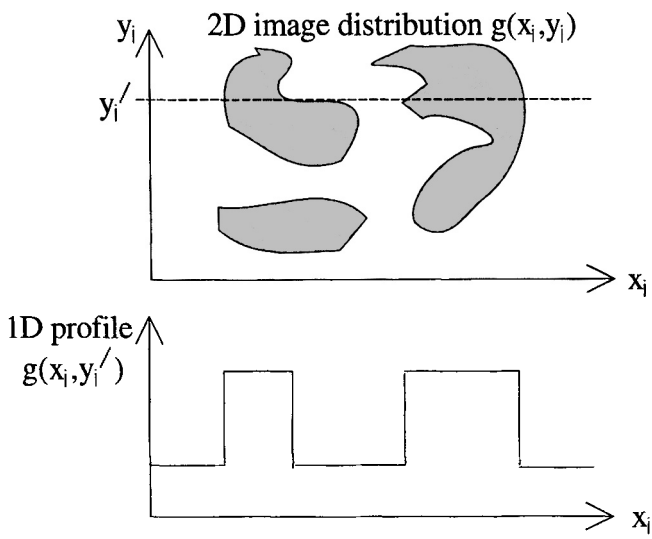


Figure 2.1. Generation of a one-dimensional waveform for Fourier analysis.

www.EngineeringEbooksPdf.com

By means of this spatial-frequency-domain description, an object or image waveform can be decomposed into its component sinusoids (Fig. 2.2), with a spatial period X (e.g., mm) and a spatial frequency $\xi \equiv 1/X$ (e.g., cycles/mm). A two-dimensional object will be decomposed into a basis set of two-dimensional sinusoids, each having two parameters. In rectangular coordinates, any given sinusoid at an arbitrary orientation has an x-domain spatial-frequency component (ξ) corresponding to the inverse of the crest-to-crest distance (X) along x, as well as a y-domain spatial-frequency component (η) corresponding to the inverse of the crest-to-crest distance (Y) along y. In polar coordinates, the same sinusoid can again be decribed by two parameters, a minimum crest-to-crest distance and an orientation angle with respect to the coordinate axes.

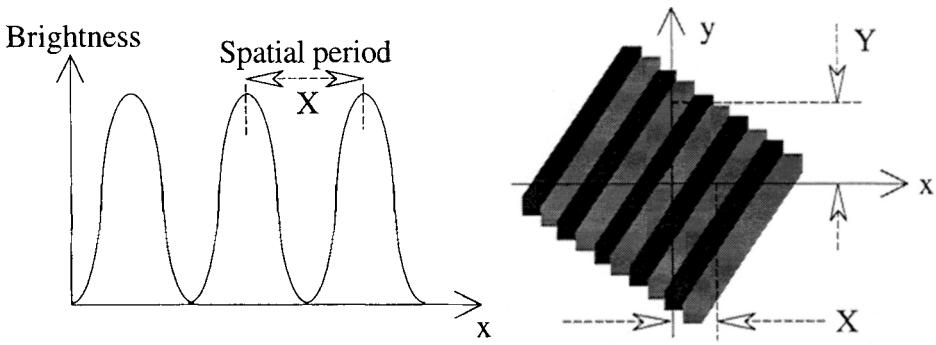


Figure 2.2. Examples of one- and two-dimensional sinusoidal basis functions.

Spatial frequency can also be described in angular units (Fig. 2.3). When the range R and spatial period X are in the same units, the angular spatial frequency $\xi_{ang} = R/X$ is in cycles/radian. However, typically the range is longer than the spatial period, yielding units for ξ_{ang} of cycles/milliradian. Angular-spatial-frequency units are often used in the specification of MTF performance for a system when the distance from the system to the object is not known exactly. For any specific situation where the distance is known, the angular units can easily be converted back to the usual units of cy/mm.

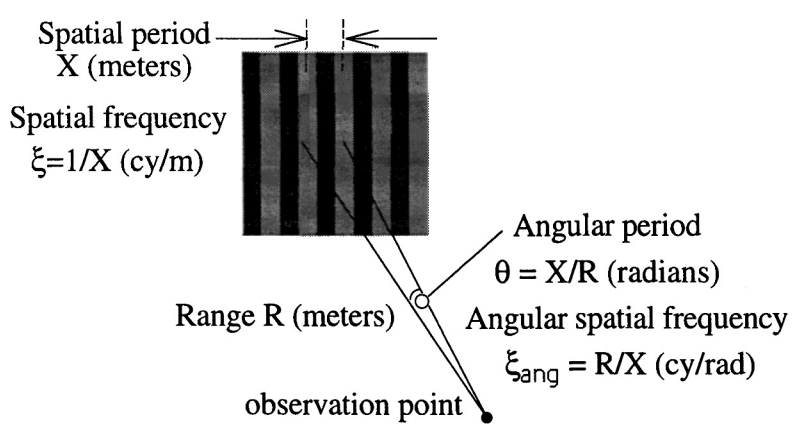


Figure 2.3. Angular spatial frequency.

www.EngineeringEbooksPdf.com

2.2 Transfer Functions

The convolution of the ideal image function with the impulse response entails a loss of fine detail. This filtering operation can be cast in the spatial-frequency domain using the convolution theorem. Taking the Fourier transform of both sides of Eq. (2.1) yields

$$F(\xi,\eta) \times H(\xi,\eta) = G(\xi,\eta) , \tag{2.2}$$

where the uppercase functions denote the Fourier transforms of the corresponding lowercase functions. We can thus interpret $F(\xi,\eta)$ as the Fourier spectrum of the object and $G(\xi,\eta)$ as the Fourier spectrum of the image. Note that the Fourier transformation has converted the convolution of Eq. (2.1) into a multiplication. Thus, $H(\xi,\eta)$, the Fourier spectrum of the impulse reponse, is a transfer function that multiplicatively relates the object and image spectra.

The frequency-domain viewpoint is convenient because multiplications are easier to perform and visualize than are convolutions. This computational and conceptual advantage is most apparent when several subsystems (e.g., atmosphere, optics, detector, electronics, display, observer) are to be cascaded, each having its own impulse response and image-quality degradation. A transfer function can be defined for each subsystem, and the resulting system transfer function can be found by a one-frequency-at-a-time multiplication of the individual transfer functions.

Usually, $H(\xi,\eta)$ is normalized such that it has value of unity at zero spatial frequency. This yields a transfer function for any given frequency that is relative to the value of the transfer function at low frequencies, and normalizes any attenuation factors that are independent of spatial frequency (such as Fresnel-reflection losses). With this normalization, $H(\xi,\eta)$ is referred to as the optical transfer function (OTF). The OTF is, in general, a complex function, having both a magnitude and a phase portion. The magnitude portion of the OTF is termed the modulation transfer function (MTF), while the phase portion is called the phase transfer function (PTF):

$$OTF(\xi,\eta) = MTF(\xi,\eta) \exp\{-jPTF(\xi,\eta)\} . \tag{2.3}$$

The interpretation of the MTF is that of the magnitude response of the optical system to sinusoids of different spatial frequencies. We consider the imaging of sinewave inputs rather than point inputs. A linear shift-invariant optical system images a sinusoid as another sinusoid. In general, the output sinusoid will have some phase shift and a reduced modulation depth, compared to the input waveform. The effect of the limited resolution of the optical system is to reduce the modulation depth in the image, compared to that in the object. We define modulation depth in Fig. 2.4 as

$$M \equiv \frac{A_{max} - A_{min}}{A_{max} + A_{min}} \tag{2.4}$$

where A_{max} and A_{min} are the maximum and minimum values of the object or image brightness in appropriate units (radiance, exitance, or irradiance). The modulation depth is the ac value of the waveform divided by the dc bias. It should be noted that, because brightness is strictly a positive quantity, $0 \le M \le 1$. We see from Eq. (2.4) that $M = 1$ whenever the minimum value of the object or image sinusoid is zero, and that $M = 0$ if the ac component of the waveform is zero.

www.EngineeringEbooksPdf.com

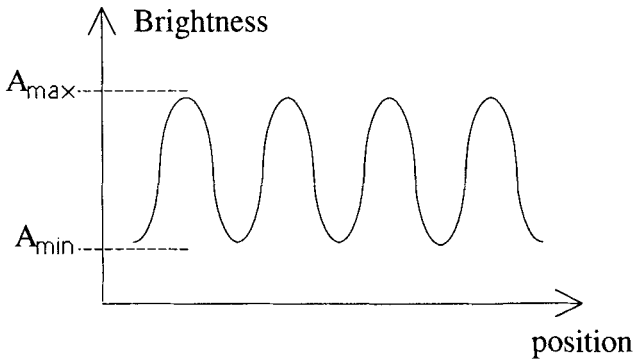


Figure 2.4. Definition of modulation depth.

Conceptually, if we test the optical system with a target consisting of a set of sinusoids of various spatial frequencies (Fig. 2.5), the MTF can be directly seen from the modulation depth in the output waveform. The image modulation depth as a function of frequency is the MTF, assuming a constant object modulation.

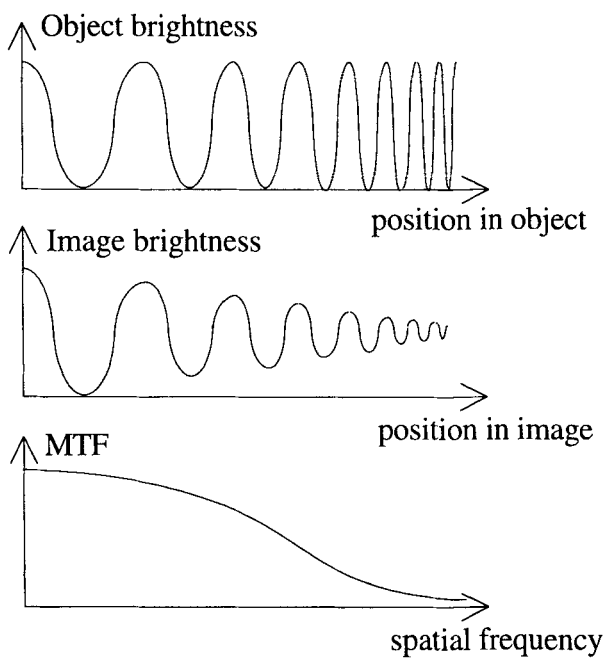


Figure 2.5. MTF is the modulation transfer as a function of spatial frequency.

Because of the limited resolution of the optical system, the valleys of the image-plane sinusoids will be less deep and the peaks of the image-plane sinusoids will be less high, decreasing the modulation depth of the image compared to the object. The modulation transfer, for any given spatial frequency ξ is the ratio of the image modulation to the object modulation. It is evident that the modulation transfer depends on spatial frequency. The limited resolution is manifest more drastically at high spatial frequencies

www.EngineeringEbooksPdf.com

where the desired detail is smaller. With increasing spatial frequency, the blurring becomes more noticeable and the modulation transfer decreases. If we plot the modulation transfer as a function of spatial frequency (assuming a constant object modulation as in Fig. 2.5) we obtain the MTF:

$$MTF(\xi) = M_{image}(\xi)/M_{object} \quad . \tag{2.5}$$

2.3 Resolution

Resolution has a number of alternate definitions. As illustrated in Fig. 2.6, it can be defined in the spatial domain as the separation in object space angle or in image plane distance for which two point sources can be discerned as being separate. It can also be defined in the spatial frequency domain as the inverse of this separation, or as the spatial frequency at which the MTF falls below a certain criterion (Fig. 2.7). When we compare the figures of merit offered by MTF and resolution, we find that resolution is usually conveniently stated as a single-number specification, while the MTF provides additional information regarding image quality over a range of spatial frequencies because it is a functional relationship rather than a single number.

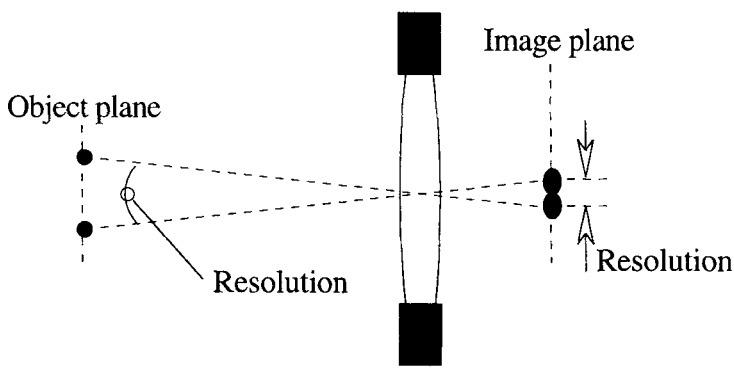


Figure 2.6. Resolution specified in terms of spacing of point sources or their images.

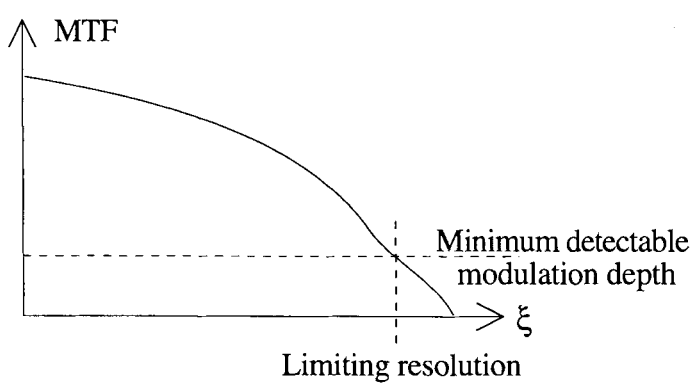


Fig. 2.7. Resolution specified in terms of a limiting spatial frequency.

www.EngineeringEbooksPdf.com

A specification of resolution alone can be at best incomplete, or at worst misleading. For example, Fig. 2.8 shows that two systems can have identical limiting resolutions, while system A clearly has superior performance over system B. In Fig. 2.9, the criterion of resolution may provide an incorrect assessment of system performance if the midband range of spatial frequencies is most important for the task at hand.

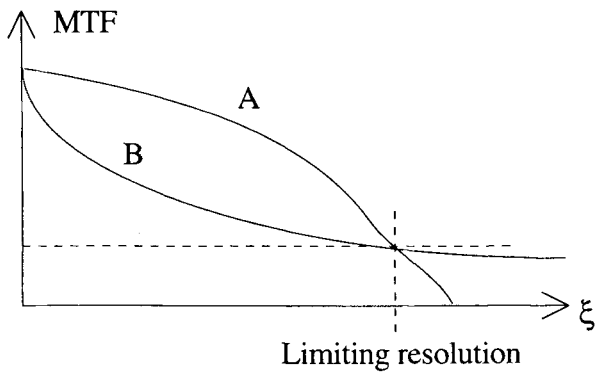


Figure 2.8. System A is better than system B, but both have identical resolution.

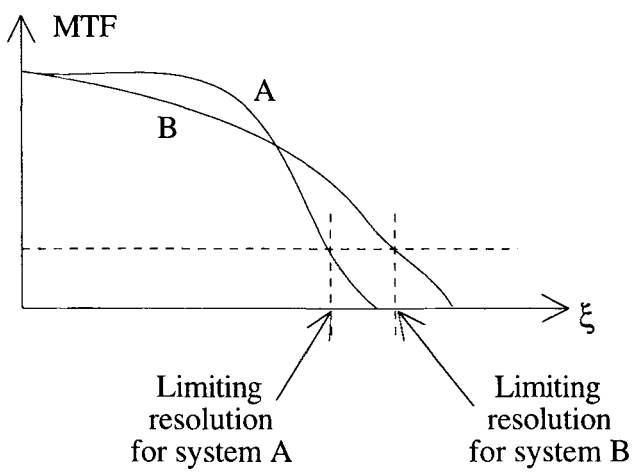


Figure 2.9. System A is better than B at midrange frequencies, but has poorer resolution.

One common criterion for image quality is to obtain the maximum area under the MTF curve, over the frequency range of interest, perhaps with a weighting function that takes into account the relative importance of a given spatial frequency band for the specific task of interest.

www.EngineeringEbooksPdf.com

2.4 MTF Calculations

We can estimate the performance of a well-corrected system by calculating the diffraction-limited MTF (Fig. 2.10). For a square aperture of dimension D × D, the MTF is of particularly simple form:

$$\mathrm{MTF}(\xi) = 1 - \xi/\xi_{\mathrm{cut}} \ , \tag{2.6}$$

where

$$\xi_{\mathrm{cut}} = 1/(\lambda F/\#) \ . \tag{2.7}$$

For a circular aperture of diameter D, we have

$$\mathrm{MTF}\left(\frac{\xi}{\xi_{\mathrm{cut}}}\right) = \frac{2}{\pi}\left\{\cos^{-1}\left(\frac{\xi}{\xi_{\mathrm{cut}}}\right) - \left(\frac{\xi}{\xi_{\mathrm{cut}}}\right)\left[1 - \left(\frac{\xi}{\xi_{\mathrm{cut}}}\right)^2\right]^{1/2}\right\}, \quad \xi \le \xi_{\mathrm{cut}} \ . \tag{2.8}$$

In Eqs. (2.7) and (2.8) we take the usual definitions of F/# as f/D for an object at infinity, as p/D for object space, and q/D for image space.

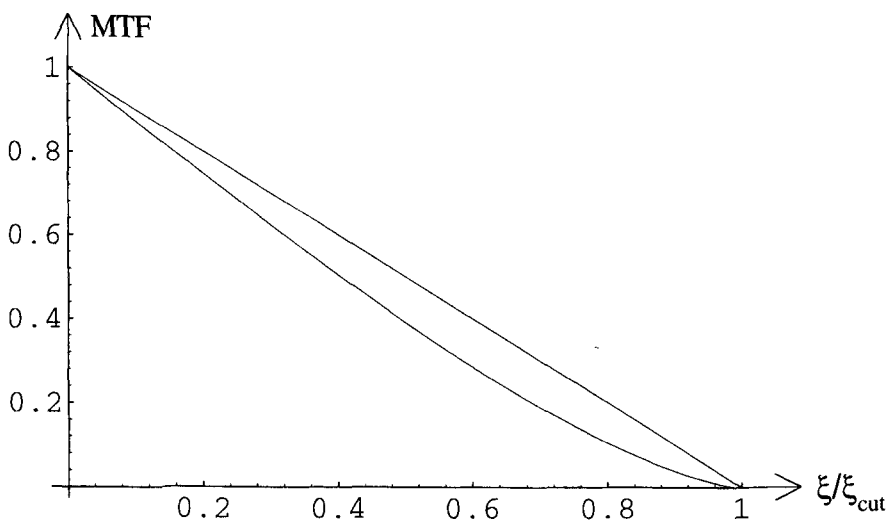


Figure 2.10. Diffraction-limited MTFs for square aperture and circular aperture.

We calculate the diffraction-limited MTF, first for the case of an object at finite conjugates and second for an object at infinity. For the finite-conjugate example, assume a diffraction-limited system with a rectangular aperture of 2 cm by 2 cm, and a wavelength of 0.6 μm. We take the object distance as p = 5 cm and the image distance as q = 10 cm. According to Eq. (2.7), the object-space cutoff frequency is $\xi_{\mathrm{cutoff,obj}}$ = [0.6 μm (5 cm/2 cm)]$^{-1}$ = 666.6 cycles/mm, and the image-space cutoff frequency is $\xi_{\mathrm{cutoff,img}}$ = [0.6 μm (10 cm/2 cm)]$^{-1}$ = 333.3 cycles/mm. The difference between object-space and image-space spatial frequencies is simply a manifestation of the difference between the object-space and image-space F/#. The same filtering caused by the finite aperture can be

www.EngineeringEbooksPdf.com

thought of as occuring in either space. These equations are useful if, for example, we want to find the spatial frequency at which the MTF is equal to 30%. According to Eq. (2.6), the MTF is 30% when the spatial frequency is 70% of the cutoff frequency, which is an image-plane spatial frequency of 233.3 cycles/mm in the image.

For an object at infinite conjugates, we find that the MTF can be expressed in terms of either angular spatial frequency in object space, or in image space in cycles/mm. As shown in the previous example, the filtering experienced in either space is equivalent. Given a focal length of 3.33 cm and a 2-cm-by-2-cm aperture stop, the cutoff angular frequency in object space is $\xi_{cut} = D/\lambda$, yielding 33.3 cycles/mrad. The corresponding spatial frequency in the image space is $1/(\lambda\ F/\#) = 1000$ cycles /mm. We verify the correspondence between angular spatial frequency in the object and the image-plane spatial frequency in Fig. 2.11, where the spatial period $X = f\ \theta = 1/\xi$. The angular spatial freqency in the object is $1/\theta$, yielding the relationship

$$\xi_{ang} = 1/\theta\ [\text{cycles/mrad}] = \xi\ [\text{cycles/mm}] \times f\ [\text{mm}] \times 0.001\ [\text{rad/mrad}] \quad . \qquad (2.9)$$

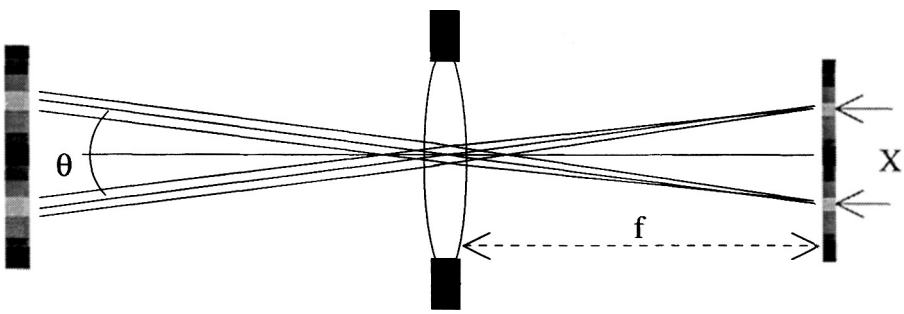


Figure 2.11. Relation of object-plane to image-plane spatial frequency.

Bibliography

Baker, L., ed., *Selected Papers on Optical Transfer Function: Foundation and Theory*, SPIE Milestone Series, vol. MS 59, Bellingham, WA, 1992.

Baker, L., ed., *Selected Papers on Optical Transfer Function: Measurement*, SPIE Milestone Series, vol. MS 60, Bellingham, WA, 1992.

Gaskill, J., *Linear Systems, Fourier Transforms, and Optics*, Wiley, New York, 1978.

Goodman, J., *Introduction to Fourier Optics*, McGraw-Hill, New York, 1968.

Williams, C., and O. Becklund, *Introduction to the Optical Transfer Function*, Wiley-Interscience, New York, 1989.

www.EngineeringEbooksPdf.com

Copyrighted Materials
Copyright © 1998 SPIE Retrieved from www.knovel.com

Chapter 3
Radiometry

3.1 Introduction

Radiometric concepts are necessary for a quantitative understanding of flux transfer through optical systems. The primary question to be addressed is: given an optical system and a source configuration, how much power from the source is collected and brought to the detector surface? The radiometric efficiency of the system is crucial to the overall signal-to-noise ratio and image detectability.

We consider the radiometry of incoherent thermal sources. Radiometry of lasers and other coherent sources is excluded. We neglect interference effects, calculating the distribution of energy over a surface as a scalar sum of powers, rather than as a vector sum of amplitudes, as would be necessary in the calculation of an interference pattern. We also neglect diffraction effects, except for the case of point sources. We will often make the approximation that an angle is small enough that its sine can be expressed as the angle itself in radians. This paraxial approach to radiometry is consistent with our treatment of geometrical optics in Chapter 1.

We use a set of SI-based units, although not all quantities are strictly MKS. We will need a number of physical constants, which are listed in Table 3.1 in the form usually used for calculations.

Table 3.1. Physical constants.

Speed of light in a vacuum	$c = 3 \times 10^{-8}$	$m\ s^{-1}$
Boltzmann's constant	$k = 1.38 \times 10^{-23}$	$J\ K^{-4}$
Planck's constant	$h = 6.6 \times 10^{-34}$	$J\ s$
Stefan-Boltzmann constant	$\sigma = 5.7 \times 10^{-12}$	$W\ cm^{-2}\ K^{-4}$

We will use the Kelvin temperature scale, which is degrees Celsius + 273. The formulae used to express radiometric quantities require an absolute temperature scale, one which does not have a zero in the middle of the range. Temperatures in the Fahrenheit or Rankine scales are also easily converted to Kelvin, as seen in Table 3.2.

Table 3.2. Temperature scales.

	Kelvin (K)	Celsius (°C)	Fahrenheit(°F)	Rankine (°R)
Absolute Zero	0	-273	-460	0
Water Freezes	273	0	32	492
Human Body	310	37	99	559
Water Boils	373	100	212	672

www.EngineeringEbooksPdf.com

3.2 Solid Angle

Solid angle is the quantity used to specify the range of pointing directions in three-dimensional space that takes one from a point in space to a surface. Solid angle is a dimensionless quantity, with units of steradian (sr). It is helpful to recall the analogous definition of radians in plane geometry seen in Fig. 3.1. An angle θ is measured in radians as the ratio of arc length s to radius r, where s and r are measured in the same units. The solid angle Ω is defined in steradians as the area on the surface of a sphere divided by the square of the radius of the sphere.

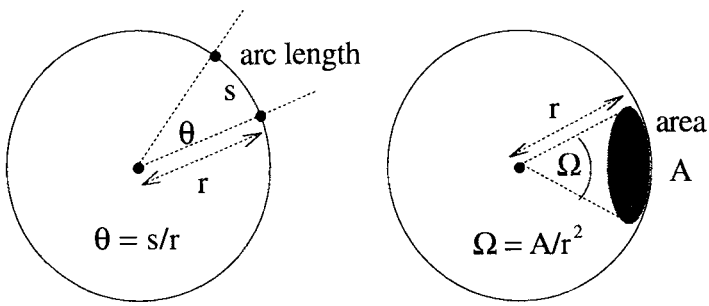


Figure 3.1. Planar angle in radians and solid angle in steradians.

Taking the surface area of a sphere as $4\pi r^2$ and the solid-angle definition

$$\Omega = A/r^2 \tag{3.1}$$

we find that a sphere contains 4π steradians of solid angle. For surface areas that are small compared to the total sphere surface, it is an acceptable approximation to simply use the area of the surface in Eq. (3.1), whether it is a flat area, or an area on the surface of the sphere. For example, a small flat disc of area A, located at a distance r, with its surface normal parallel to the line of sight has a solid-angle subtense of $\Omega = A/r^2$. A flat disc is a good approximation to the entrance aperture of an optical system, and this result holds for other flat surfaces, such as squares, as long as the linear dimension of the surface is small compared to the viewing distance. A more exact derivation of the solid angle for a flat disc yields the expression

$$\Omega_{disc} = 4\pi \sin^2(\varphi/2) \tag{3.2}$$

where φ is the planar half-angle of the disc as seen in Fig. 3.2. It is easily shown that this expression goes to A/r^2 for small angles φ.

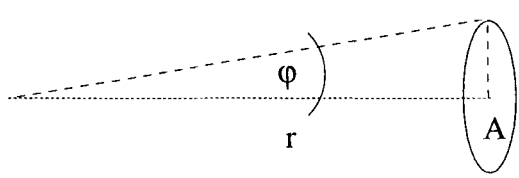


Figure 3.2. Definition of φ for small flat disc.

www.EngineeringEbooksPdf.com

Another useful expression is the solid angle of a small flat area which is tilted with respect to the line of sight. The surface normal makes an angle θ with the line of sight, as seen in Fig. 3.3. The solid angle is decreased by the factor cosθ to yield

$$\Omega = A/r^2 \cos\theta \ . \tag{3.3}$$

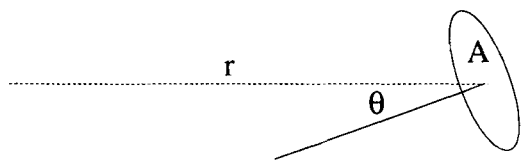


Figure 3.3. Solid angle of a tilted flat disc.

3.3 Flux-Transfer Calculations

Quantitative characterization of flux transfer requires its own set of units. Table 3.3 shows the pertinent quantities. Calculational examples of each will follow.

Table 3.3. Energy-based radiometric units.

Symbol	Quantity	Units
Q	Energy	J
ϕ	Flux (power)	W = J/s
I	Intensity	W/sr
E	Irradiance	W/cm^2
M	Exitance	W/cm^2
L	Radiance	W/(cm^2 sr)

These units are derived from a number of Joules, and are referred to as energy-based units. An analogous set of units can be derived on the basis of a number of photons. Photon-derived quantities (Table 3.4) are denoted with a subscript q.

Table 3.4. Photon-based radiometric units.

Symbol	Quantity	Units
Q_q	Photon number	Photons
ϕ_q	Photon flux	Photons/s
I_q	Photon intensity	Photons/(s sr)
E_q	Photon irradiance	Photons/(s cm^2)
M_q	Photon exitance	Photons/(s cm^2)
L_q	Photon radiance	Photons/(s cm^2 sr)

www.EngineeringEbooksPdf.com

Conversion between the two sets of units is easily accomplished using the relationship for the amount of energy carried per photon

$$\mathscr{E} = hc/\lambda \ . \tag{3.4}$$

A units analysis yields, for example, ϕ (J/s) = ϕ_q (photon/s) × $\mathscr{E}$ (J/photon). We will see in Chapter 4 that the photon-based units are useful in the description of detectors that respond directly to absorbed photons (so-called photon detectors), while the energy-based units are appropriate for the description of detectors which respond to absorbed thermal energy (so-called thermal detectors). Note that $\mathscr{E}$, and thus the conversion between the two sets of units, depends on λ. A longer-wavelength photon carries less energy than a short-wavelength photon. At an infrared wavelength of 10 μm, the photon energy is approximately 2×10^{-20} J/photon, while at a visible wavelength of 0.5 μm, the photons are a factor of 20 more energetic, having approximately 4×10^{-19} J/photon. This units conversion can also be thought of in terms of λ/hc, that is, how many photons/s it takes to make 1 W of power. At 10 μm, there are 5×10^{19} photons/s in 1 W. For the more energetic photons at a shorter wavelength of 0.5 μm, fewer photons/s (2.5×10^{18}) are required to make 1 W. Note that when wavelength-dependent quantities are plotted, the use of either of these two different sets of units will yield curves of different shape.

Irradiance and exitance have the same units (W/cm^2), but a different interpretation. Irradiance characterizes the spatial power density associated with a surface that receives power, while exitance refers to a surface that emits power. Referring to Fig. 3.4, we assume that a uniform irradiance of 4 W/cm^2 is incident on a 1-cm-by-1-cm surface. Consider a small portion of that surface as a single detector element in a two-dimensional focal-plane array of detectors. How much power falls onto a 1-mm-by-1-mm portion of the surface? We use the relationship

$$\phi = A \times E \tag{3.5}$$

which yields a flux received of 40 mW. If a uniform irradiance is specified, the flux received by a detector element is, according to Eq. (3.5), simply the irradiance multiplied by the area of the detector, assuming that the detector is smaller than the irradiance distribution.

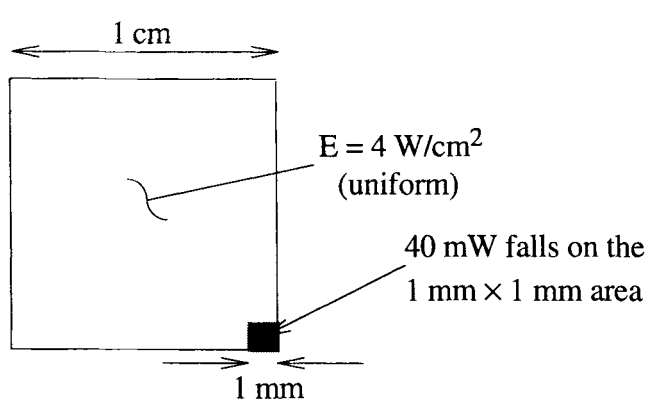


Figure 3.4. Irradiance example.

www.EngineeringEbooksPdf.com

Exitance is used to calculate the total power radiated from an emitting surface. For example, let the radiating surface have an emitting area (1 mm)^2, and a uniform exitance of 3 W/cm^2. The total power radiated from that surface is seen to be 30 mW, calculated from

$$\phi = A \times M \quad . \tag{3.6}$$

Note that the power ϕ in Eq. (3.6) is the total power that would be collected by a hemisphere that bounds the radiating surface, as seen in Fig. 2.5. The distribution of irradiance over the surface of the hemisphere is not uniform, as we will see in the description of radiance.

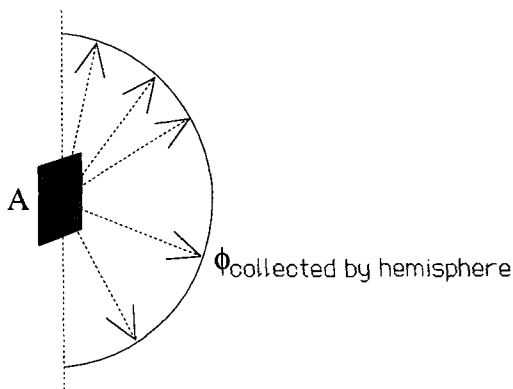


Figure 3.5. Exitance example.

Intensity is measured in W/sr, and it is the quantity that must be used to characterize radiation from a point source. Intensity can also be used to characterize extended sources. One definition of a point source is that the source has a linear dimension that is small compared to the viewing distance. A point source can also be defined as a source with a dimension that is smaller than the resolution element of the optical system (the source is smaller than the projection of the resolution spot at the object plane). The simplest calculation involving intensity is: how many watts are radiated by a point source with a uniform intensity? The total solid angle into which the point source radiates is 4π, so for an intensity of 3 W/sr, we find, using

$$\phi = I \times \Omega \ , \tag{3.7}$$

that the total radiated power is 12 π W. This example does not represent the typical calculation, because we do not usually collect the flux from a point source over all directions, but rather we are concerned with the flux collected by a receiver (either a detector or the entrance aperture of an optical system) having a given solid angle subtense when seen from the point source. Thus, the calculation becomes a two-step process: first we calculate the solid angle of the recevier as seen from the point source, and then we use Eq. (3.7) to calculate the flux transferred. For example, assume we want to find the power transmitted by a circular aperture in an opaque screen, as seen in Fig. 3.6, for the

www.EngineeringEbooksPdf.com

source discussed above. The aperture has a radius of 1 mm, and the screen is placed at a distance of 3 cm from the point source. First we calculate the solid angle of the aperture as seen from the point source using Eq. (3.1). This yields a solid angle of $\Omega = A/r^2 = 3.5 \times 10^{-3}$ sr. Next, using Eq. (3.7) we calculate $\phi = I \times \Omega = 10.5$ mW.

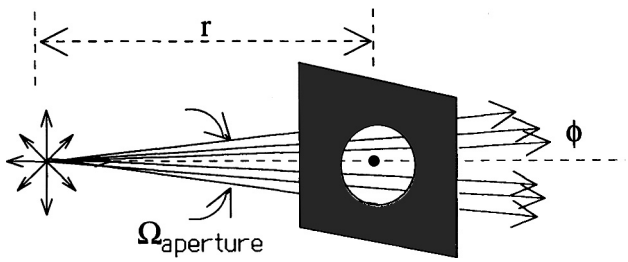
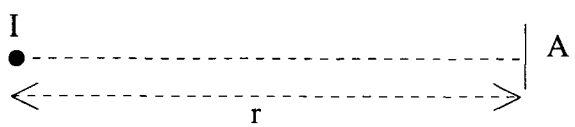


Figure 3.6. Intensity example.

Another example of intensity calculations is that irradiance falls off as $1/r^2$ from a point source, as seen in Fig. 3.7. Consider a receiver of area A, which is placed at various distances from a point source of intensity I. Both the flux and the irradiance fall off as $1/r^2$, which can be thought of as the condition that requires the same amount of flux to pass through any particular sphere. For calculation of irradiance at any location, the flux collected on a small area a is determined, and divided by that area. This procedure provides a convenient means to solve for E, particularly in more complex geometries, such as seen in Fig. 3.8, where we solve for E as a function of θ on a flat screen placed at a distance r from a point source.



$$\phi = I \times \Omega = I \times A/r^2$$
$$E = \phi/A = I/r^2$$

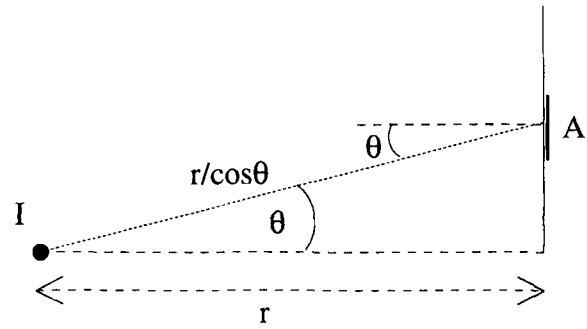
Figure 3.7. Irradiance falls off as $1/r^2$ from a point source.

Radiance, L, has units of $W/(cm^2\ sr)$ and is used to characterize extended sources. Radiance is defined for a particular ray direction, as seen in Fig. 3.9, as the radiant power per unit projected source area (perpendicular to the ray) per unit solid angle. We see that

$$L = \frac{\partial^2 \phi}{\partial A_s \cos\theta_s\, \partial \Omega_d}, \tag{3.8}$$

where $\partial^2\phi$ is the power radiated into the cone. The double differential indicates that $\partial^2\phi$ is incremental with respect to both projected area and solid angle.

www.EngineeringEbooksPdf.com



$$\phi = I \times \Omega = I \times (A \cos\theta)/(r/\cos\theta)^2$$
$$\phi = I\, A\, (\cos^3\theta) / (r^2)$$
$$E = \phi/A = I\, (\cos^3\theta) / (r^2)$$

Figure 3.8. An off-axis geometry for irradiance calculation.

Cross multiplying Eq. (3.8), we obtain

$$\partial^2\phi = L\, \partial A_s \cos\theta_s\, \partial\Omega_d \quad . \tag{3.9}$$

Note that because of the definition of radiance in terms of projected source area, $\partial^2\phi$ has a cosine-θ dependence even for the case of a Lambertian radiator where L is a constant, independent of θ. An approximation can be made for the case of small but finite-sized source area A_s and receiver solid angle Ω_d

$$\phi \approx L\, A_s \cos\theta_s\, \Omega_d \quad . \tag{3.10}$$

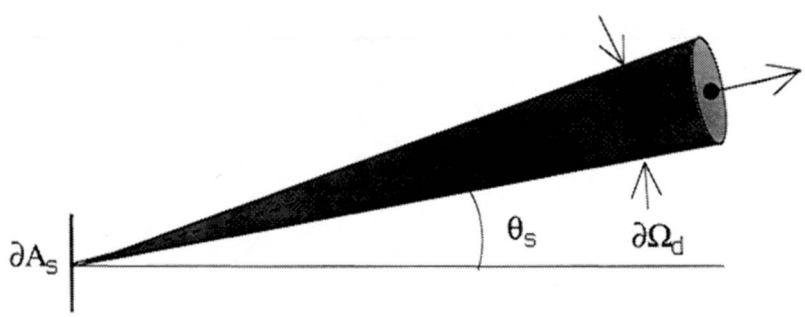


Figure 3.9. Geometry for definition of radiance.

For the simplest calculational example of radiance, let the view angle $\theta_s = 0$, as seen in Fig. 3.10. Given an extended source of area 1 cm by 1 cm, specified by a radiance of 5 W/(cm^2 sr). At an axial distance of 10 m, there is located a detector of dimensions 1 mm by 1 mm. How much power falls on the detector?

www.EngineeringEbooksPdf.com

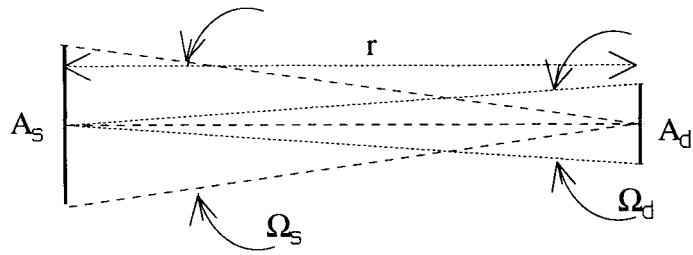


Figure 3.10. Geometry for flux-transfer calculation using radiance.

First, we calculate the solid angle of the detector as seen from the source, $\Omega_d = A_d/r^2 = 10^{-8}$ sr. We multiply this solid angle by the area of the source and the radiance of the source to obtain the power on the detector: $\phi_{det} = L\, A_s\, \Omega_d = 5 \times 10^{-8}$ W. We can write the equation for the flux on the detector in two equivalent alternate ways, depending on whether the r^2 is grouped with A_s or A_d. Assuming small angles,

$$\phi = L\, A_s\, \Omega_d = L\, \frac{A_s\, A_d}{r^2} = L\, A_d\, \Omega_s \quad . \tag{3.11}$$

The flux on the detector is equal to the radiance of the source multiplied by an AΩ product, either the area of the source and the solid angle of the detector, or the area of the detector and the solid angle of the source. We can take whichever AΩ product is most convenient because they are equivalent, within the small-angle approximation. The AΩ product can be thought of as a throughput or flux-gathering capability. The AΩ product is a way of counting up how many rays get from source to receiver, as seen in Fig. 3.11. If we assume that each ray carries a certain incremental amount of power, AΩ product is proportional to flux transferred.

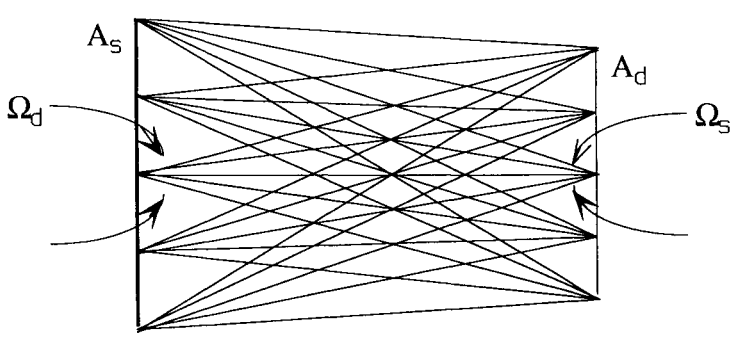


Figure 3.11. AΩ product for free-space propagation.

Using our expression Eq. (3.3) for the solid angle of a tilted receiver, we can write the following expression for the case of flux transfer to an on-axis tilted detector, seen in Fig. 3.12. If we define the angle θ_d to be the angle between the line of centers

www.EngineeringEbooksPdf.com

and the surface normal to the detector, we see in Eqs. (3.12) and (3.13) that both the flux collected on the detector and irradiance on the detector are decreased by a factor of $\cos\theta_d$:

$$\phi = L\, A_s\, \Omega_d = L\, A_s\, \frac{A_d \cos\theta_d}{r^2} \tag{3.12}$$

$$E = \frac{\phi}{A_d} = \frac{L\, A_s \cos\theta_d}{r^2} \quad . \tag{3.13}$$

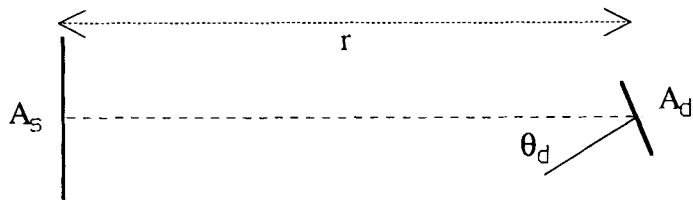


Figure 3.12. Flux transfer to tilted receiver.

Before we consider flux transfer in situations where $\theta_s \neq 0$, we need to define the concept of a lambertian radiator, which has a radiance L that is independent of view angle θ_s. This does not mean that a lambertian radiator radiates an equal amount of flux into all solid angles. Referring to Eq. (3.9) we find that $\partial^2\phi$ (the amount of flux radiated into a cone of solid angle $\partial\Omega$ from an area ∂A_s) falls off proportional to $\cos\theta_s$, for the case where L is independent of θ_s. For a nonlambertian radiator, L is explicitly a function of θ_s, so generally $\partial^2\phi$ falls off faster than $\cos\theta_s$.

For a lambertian source, the following relation holds between L and M:

$$M\ [\mathrm{W/cm^2}] = L\ [\mathrm{W/(cm^2\ sr)}] \times \pi\ [\mathrm{sr}] \quad . \tag{3.14}$$

The units of this equation are consistent, but since the flat lambertian source radiates into the full forward hemisphere, why is the proportionality factor equal to π and not 2π? The flux $\partial^2\phi$ from the lambertian source falls off proportional to $\cos\theta_s$, yielding a nonuniform illumination of the hemisphere (Fig. 3.13). An integration over all θ_s in the hemisphere yields a factor of π rather than a factor of 2π. We begin with the definition of exitance:

$$\phi_{\text{hemisphere}} = M\, A_s \quad . \tag{3.15}$$

Then we define the flux reaching the hemisphere from the lambertian source as a sum:

$$\phi_{\text{hemisphere}} = \sum_{\cos\theta_s=1}^{0} \partial^2\phi \quad . \tag{3.16}$$

Expressing this sum as an integral:

www.EngineeringEbooksPdf.com

$$\phi_{hemisphere} = \int_{hemisphere} L\, A_s \cos\theta_s \, d\Omega_d \quad , \tag{3.17}$$

where in spherical coordinates,

$$d\Omega_d = \frac{dA_d}{r^2} = \frac{r^2 \sin\theta \, d\theta \, d\varphi}{r^2} = \sin\theta \, d\theta \, d\varphi \quad . \tag{3.18}$$

Substituting Eq. (3.18) into Eq. (3.17), we find that

$$\phi_{hemisphere} = \int_{-\pi}^{\pi} \int_{-\pi}^{-\pi} L\, A_s \cos\theta_s \sin\theta \, d\theta \, d\varphi \quad . \tag{3.19}$$

Evaluating the integral in Eq. (3.19) yields:

$$\phi_{hemisphere} = \pi\, L\, A_s \quad . \tag{3.20}$$

Substitution of Eq. (3.20) into Eq. (3.15) yields the desired relationship between exitance and radiance for a lambertian source, Eq. (3.14).

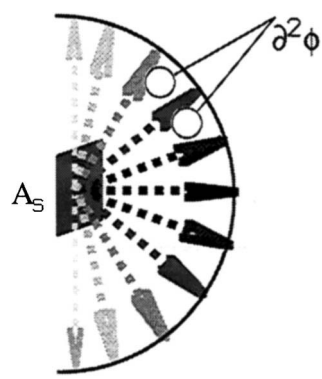


Figure 3.13. Relationship between exitance and radiance for a lambertian radiator.

We now consider flux transfer from flat lambertian sources having nonzero view angles θ_s. Considering the case where $\theta_s \neq 0$ and $\theta_d = 0$, we find a $\cos^3\theta_s$ dependence of flux received by the detector, as seen in Fig. 3.14. Compared to the perpendicular distance r, the longer line-of-centers distance $\mathcal{r} = r/\cos\theta_s$ is used in the calculation of the solid angle of the detector. Combined with the $\cos\theta_s$ flux falloff inherent in a lambertian source, this yields a cosine-cubed dependence

www.EngineeringEbooksPdf.com

$$\phi_d = L\, A_s \cos\theta_s\, \Omega_d = L\, A_s \cos\theta_s \frac{A_d}{r^2} \tag{3.21}$$

$$\phi_d = L\, A_s \cos\theta_s \frac{A_d}{(r/\cos\theta_s)^2} \propto \cos^3\theta_s \tag{3.22}$$

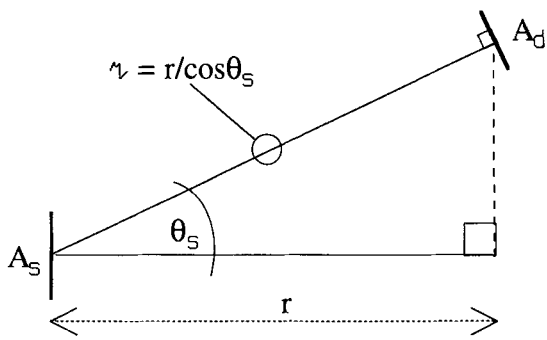


Figure 3.14. Flux transfer to an off-axis detector.

Our final example of the use of radiance is the flux transfer between two parallel surfaces, as seen in Fig. 3.15. This leads to the well-known cosine-to-the-fourth falloff in flux. For this geometry, the flux transfer equation becomes

$$\phi_d = L\, A_s \cos\theta_s\, \Omega_d = L\, A_s \cos\theta_s \frac{A_d \cos\theta_d}{r^2} \tag{3.23}$$

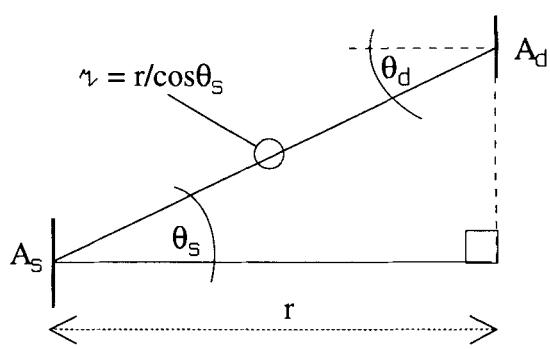


Figure 3.15. Geometry for cosine-to-the-fourth falloff.

If we use the fact that the source and detector are parallel and set $\theta_s = \theta_d = \theta$, we find the $\cos^4\theta$ dependence:

www.EngineeringEbooksPdf.com

$$\phi_d = L\, A_s \cos\theta_s \frac{A_d \cos\theta_d}{\left(r/\cos\theta_s\right)^2} \propto \cos^4\theta \quad . \tag{3.24}$$

We now consider paraxial flux transfer in image-forming systems. The calculation is performed in two steps. The first is to evaluate the flux collected by the optical system and transferred to the image plane. This is the standard $\phi = L \times A\Omega$ equation between the source and the collecting aperture of the lens. Referring to Fig. 3.16, we can see that this equation (neglecting reflection and scattering losses in the lens) can take any one of the following forms because of the equivalence of the $A\Omega$ product:

$$\phi_{collected} = L_{obj}\, A_{obj}\, \Omega_{lens\ from\ obj} \tag{3.25}$$

$$\phi_{collected} = L_{obj}\, A_{img}\, \Omega_{lens\ from\ img} \tag{3.26}$$

$$\phi_{collected} = L_{obj}\, A_{lens}\, \Omega_{obj} \tag{3.27}$$

$$\phi_{collected} = L_{obj}\, A_{lens}\, \Omega_{img} \quad . \tag{3.28}$$

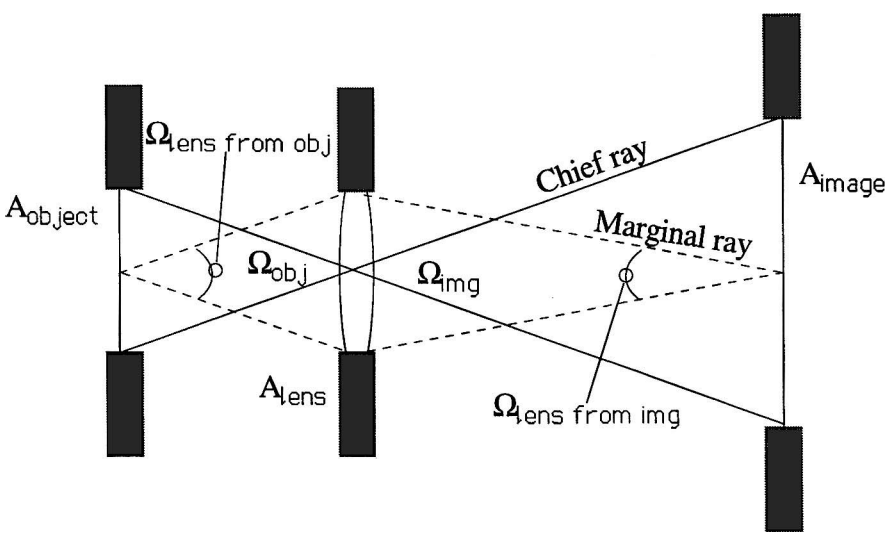


Figure 3.16. Flux transfer in an image-forming system.

The $A\Omega$ products can be interpreted as counting the ray paths from object to image, as seen in Fig. 3.17. The total flux transferred from object to image is proportional to the FOV^2 and $[1/(F/\#)]^2$. Recall that FOV and F/# are defined in terms of planar angles, so that areas and solid angles are the squares of these quantities.

The second part of the calculation for flux transfer in an image-forming system is to find the image irradiance. This is a quantity of interest from the point of view of evaluating the exposure of a detector array or a piece of film. For an image-plane flux distribution that is larger in extent than the size of an individual image-receiver element (pixel or film grain), the energy received is the power incident on the element multiplied

www.EngineeringEbooksPdf.com

by the exposure time. The power incident on the receiving element is the image-plane irradiance multiplied by the area of the individual element, according to Eq. (3.5). The image-plane irradiance is the flux collected (found using Eqs. (3.25) through (3.28)) divided by the illuminated image area. Using Eq. (1.7) for the area magnification, we find the image area is $A_{obj} \times (q/p)^2$.

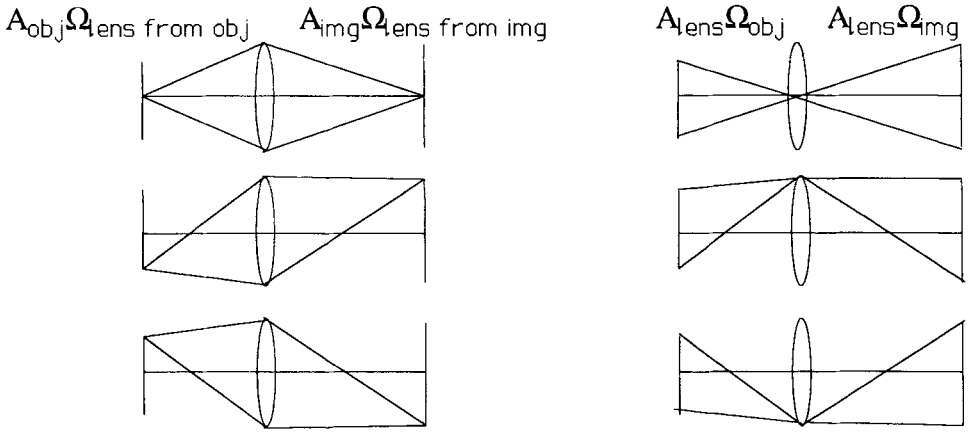


Figure 3.17. Equivalent AΩ products for paraxial image-forming systems.

Carrying out this calculation yields a significant result. The image-plane irradiance for an extended-source object is seen to depend only on the image-space F/# of the imaging system. Beginning with Eq. (3.26), and dividing both sides of the equation by A_{img}, we obtain (in the paraxial, or large-F/#, approximation)

$$E_{image} = L_{obj}\,\Omega_{lens\ from\ img} = L_{obj} \times \pi/4 \times [1/(F/\#)]^2 \ . \qquad (3.29)$$

For imaging systems where the chief ray makes a substantial angle with the optic axis, the image-plane irradiance will have a cosine-to-the-fourth falloff consistent with Eq. (3.24). Equation (3.28) implies that the amount of flux reaching the image plane is numerically equivalent to having a source the size of the lens, which has the same radiance as the original object. The lens is considered as a flat surface with a given area and radiance. The flux transfer to the image plane produces the falloff in irradiance seen in Fig. 3.18.

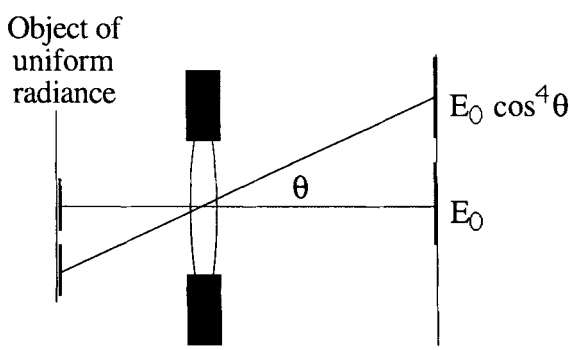


Figure 3.18. Cosine-to-the-fourth falloff in image plane irradiance.

www.EngineeringEbooksPdf.com

Example of an irradiance calculation

Consider the system seen in Fig. 3.19, with the following parameters:

$p = 5$ cm; $q = 3$ cm; $M = -q/p = -3/5$; $A_{lens} = 1$ cm^2; $L_{obj} = 2\text{W}/(\text{cm}^2\ \text{sr})$.

The object is 1 mm × 1 mm. $A_{obj} = 0.01$ cm^2 and $A_{img} = A_{obj} \times \mathcal{M}^2 = 0.0036$ cm^2.

Using Eq. (3.25), $\phi_{collected} = L_{obj}\, A_{obj}\, \Omega_{lens\ from\ obj} = L_{obj}\, A_{obj}\, (A_{lens}/p^2) = 8 \times 10^{-4}$ W.

On-axis image irradiance, $E = \phi_{collected}/A_{img} = 8 \times 10^{-4}$ W/0.0036 cm^2 = 0.22W/cm^2.

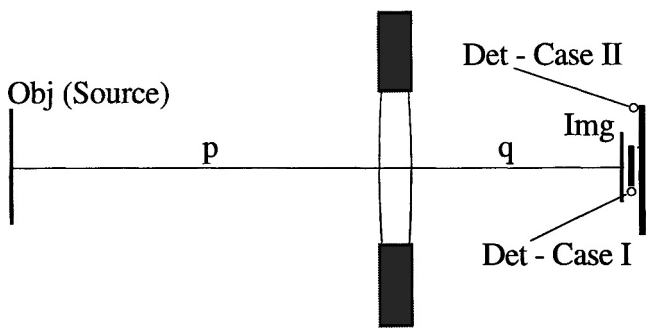


Figure 3.19. Image irradiance calculation example.

Using image irradiance we can calculate the flux on a detector, provided that the detector is smaller than the illuminated region of the image. This is the situation in Fig. 3.19, Case I, where the detector is smaller than the image of the source, and hence acts as the field stop. The flux collected by this detector is consistent with Eq. (3.5), $\phi_{det} = E_{image} \times A_{det}$. A 20-μm-by-20-μm detector will thus collect 0.88 μW of power. Equation (3.5) is not valid for Fig. 3.19 (Case II). If the detector is larger than the image of the source, the source acts as the field stop, and the detector receives all the flux transferred to the image plane. For the example given, the detector receives 8×10^{-4} W.

We have thus far considered extended-source imaging. The radiometry for point-source imaging is affected by the finite size of the impulse response formed by the optical system, as seen in Fig. 3.20. Assuming that the system is diffraction-limited, 84% of the flux collected by the lens from the point source, $\phi = I\, \Omega_{lens}$, is concentrated into a blur spot of diameter $d = 2.44\ \lambda\ (\text{F/\#})_{image\text{-}space}$. The average image-plane irradiance (a flat-top approximation within the first lobe of the diffraction pattern) is

$$E_{avg} = 0.84\, I_{source} \left(\frac{\frac{\pi}{4} D_{lens}^2}{p^2} \right) \frac{1}{\frac{\pi}{4} \left(2.44\ \lambda\ \frac{q}{D_{lens}} \right)^2} . \qquad (3.30)$$

If the area of the optic is held constant, the flux collected is a constant. If the F/# is decreased, while the system is kept in a diffraction-limited state of correction, the irradiance will increase because of the smaller image area.

www.EngineeringEbooksPdf.com

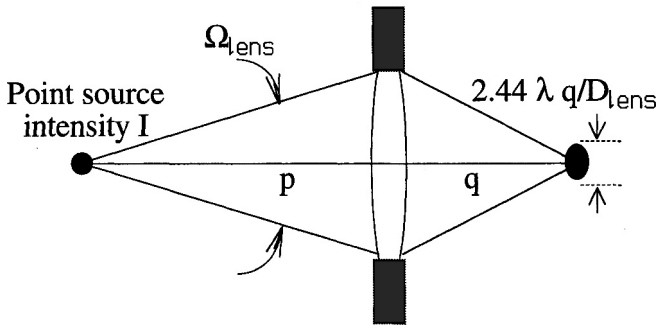


Figure 3.20. Radiometry for point-source imaging.

Bibliography

Boyd, R., *Radiometry and the Detection of Optical Radiation*, Wiley, New York, 1984.

Dereniak, E., and G. Boreman, *Infrared Detectors and Systems*, Wiley, New York, 1996.

Wolfe, W., and G. Zissis, *The Infrared Handbook*, Environmental Research Institute of Michigan, Ann Arbor, 1978.

www.EngineeringEbooksPdf.com

Copyrighted Materials
Copyright © 1998 SPIE Retrieved from www.knovel.com

Chapter 4
Sources of Radiation

4.1 Introduction

Heated solid bodies emit radiation that is particularly concentrated in the infrared region of the spectrum, roughly from 1 to 10 μm in wavelength. Such heated solid bodies emit their radiation in a continuum of wavelengths, rather than at a collection of discrete spectral lines characteristic of gaseous emission. To describe the radiation characteristic of a source which emits a finite total power over a range of wavelengths, we must introduce spectral quantities. Spectral quantities have units of μm in the denominator, and are denoted with a subscript λ to distinguish them from the corresponding integrated quantities. For example, spectral radiance L_λ has units of W/(cm^2 sr μm), and is the radiance per μm of wavelength interval. For a measurement over a 1-μm spectral bandpass, the integrated radiance (also called the in-band radiance) numerically equals the spectral radiance. For spectral bandpasses larger or smaller than 1 μm, the integrated radiance scales with $\Delta\lambda \equiv \lambda_1 - \lambda_2$, as seen in Eqs. (4.1) and (4.2):

$$L\left[\frac{\text{Watt}}{\text{cm}^2\ \text{sr}}\right] = \int_{\lambda_1}^{\lambda_2} L_\lambda\left[\frac{\text{Watt}}{\text{cm}^2\ \text{sr}\ \mu\text{m}}\right] d\lambda\,[\mu\text{m}] \tag{4.1}$$

$$M\left[\frac{\text{Watt}}{\text{cm}^2}\right] = \int_{\lambda_1}^{\lambda_2} M_\lambda\left[\frac{\text{Watt}}{\text{cm}^2\ \mu\text{m}}\right] d\lambda\,[\mu\text{m}] \tag{4.2}$$

The integration is usually carried out numerically, using the trapezoidal approximation seen in Fig. 4.1. This process can be made as accurate as desired, by choice of the interval size Δλ. As seen in Fig. 4.2, to maintain accuracy for a wider passband, individual intervals are summed together. It is necessary to choose interval locations properly if the desired passband includes a peak of the radiation curve.

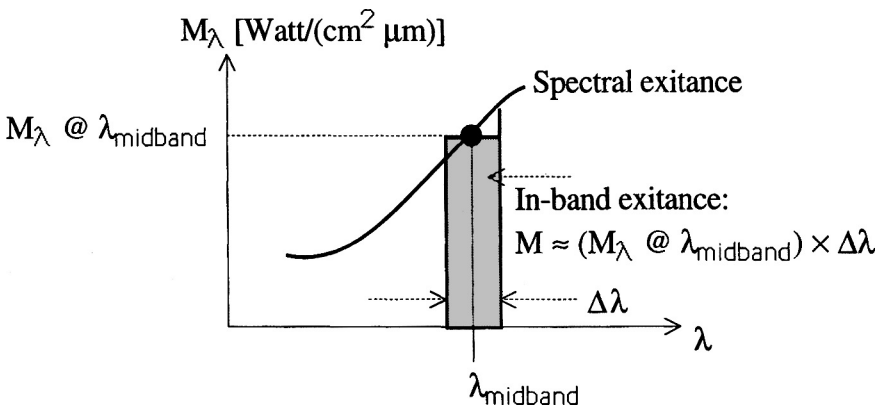


Figure 4.1. Trapezoidal integration.

www.EngineeringEbooksPdf.com

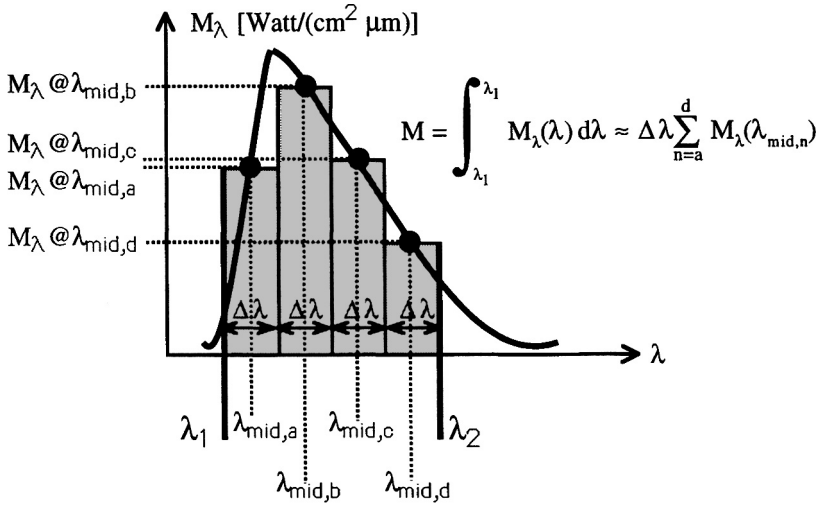


Figure 4.2. Integration over a wider passband.

4.2 Blackbody Radiation

A blackbody is an ideal radiation source, which can be completely described by specifying the temperature. A blackbody has the maximum spectral exitance possible for a heated body at a particular specified temperature. This condition is true over any $\Delta\lambda$ interval, and also over all wavelengths. A blackbody is a convenient baseline for comparison with real sources, because any real source at a given temperature is constrained to emit less radiation than the blackbody source at that temperature. Because of the population inversion inherent in laser gain media, the usual thermodynamic definition of temperature does not apply. Lasers can have spectral radiances that far exceed any blackbody source, because of their narrowband emission and their narrow beamwidth.

Blackbody radiation is also called cavity radiation. The radiation emitted from a small aperture in front of an isothermal cavity depends only on the temperature of the enclosure and not on the material of which the cavity is constructed. Typical cavity configurations are seen in Fig. 4.3, which are the basis for construction of practical blackbody simulators.

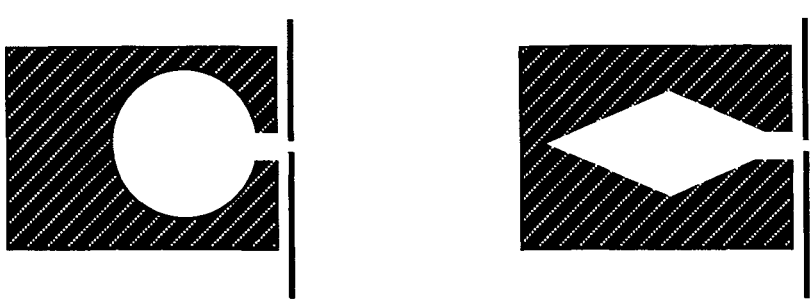


Figure 4.3. Cavity-radiator configurations.

www.EngineeringEbooksPdf.com

The spectral exitance of a blackbody is a function of two variables: temperature and wavelength. For a given temperature, exitance is a function of wavelength. This results in a family of spectral exitance curves dependent on temperature. The equation for spectral exitance is the Planck equation:

$$M_\lambda(\lambda,T) = \frac{2\pi h c^2}{\lambda^5 \left(\exp\{hc/\lambda kT\} - 1\right)} . \tag{4.3}$$

Blackbodies are lambertian radiators by definition, so their spectral radiance is simply their spectral exitance function divided by π, consistent with Eq. (3.15). Figures 4.4 and 4.5 show a family of Planck curves, covering a range of temperatures from 600 K to 2000 K. Two characteristics can be noted directly from the plots of the Planck equations. As the source temperature increases, the total area under the curve (integrated exitance) increases much faster than the absolute temperature. We will quantify this observation as the Stefan-Boltzmann Law. As the source temperature increases, the wavelength of peak exitance moves toward shorter wavelengths. This relationship is known as the Wien Displacement Law.

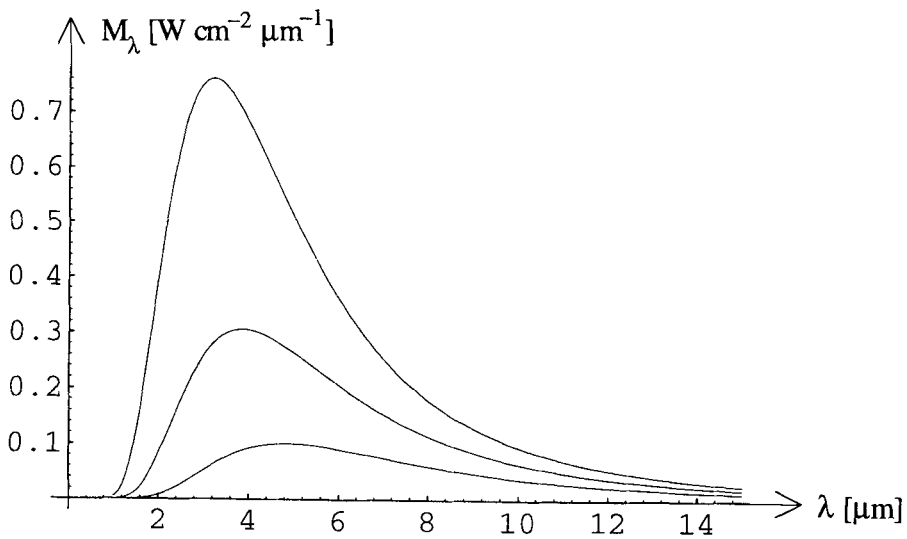


Figure 4.4. Blackbody spectral exitance for temperatures of 600, 750, and 900 K.

These plots can be used directly for approximate calculation of in-band quantities, as in the example of a blackbody source at 800 K with an area of 1 cm^2. A 1-mm × 1-mm detector is placed at a range of 1 m. How much power falls on the detector within the 8- to 10-μm band? We must calculate the radiance L that goes into the usual flux-transfer equation $\phi = LA_s\Omega_d$. Because the source in question is a blackbody, we can use Eq. (3.14), $L = M/\pi$. The quantity M is the integrated in-band exitance, which we find by a trapezoidal integration. Assuming that an integration using one trapezoid is sufficiently accurate, $M = M_\lambda(\lambda_{midband})\Delta\lambda$. From the plot in Fig. 4.4, we find that $M_\lambda(\lambda = 9\mu m) = 0.1$ W cm^{-2} μm^{-1}. Thus, with a $\Delta\lambda = 2$ μm, $M = 0.2$ W/cm^2. Hence, the in-band flux on the detector $\phi = 6.4 \times 10^{-8}$ W.

www.EngineeringEbooksPdf.com

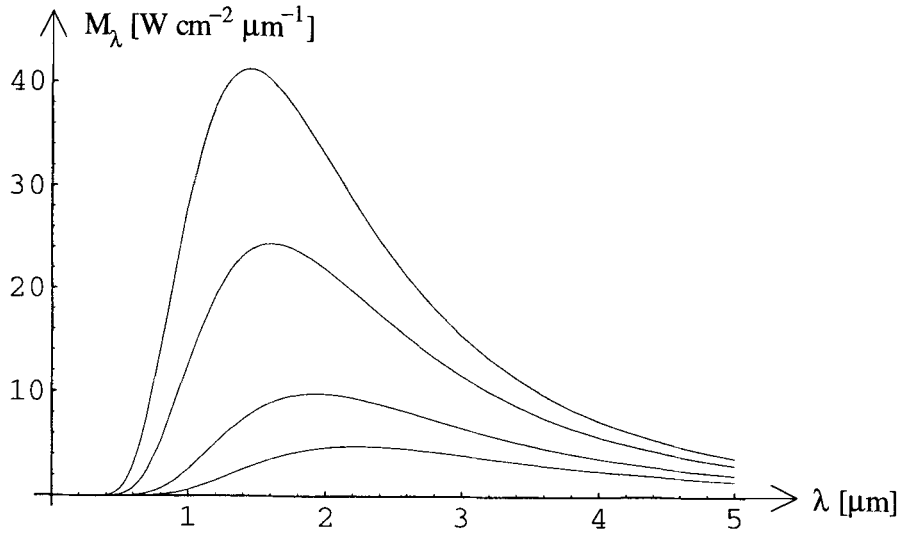


Figure 4.5. Blackbody spectral exitance for temperatures of 1300, 1500, 1800, 2000 K.

For other temperatures or other wavelength ranges, or to obtain more accurate numerical values, calculations can be made directly from Eq. (4.3). Calculation algorithms should be verified using the plots in Fig. 4.4 and 4.5, taking care to ensure that the correct units (W cm^{-2} μm^{-1}) are obtained. As an example, we calculate $M_\lambda(\lambda = 3$ μm, T = 900 K). Checking the graph of Fig. 4.4, we expect that our calculation will yield approximately $M_\lambda = 0.75$ W cm^{-2} μm^{-1}.

Considering the units in the expression of Eq. (4.3), hc/λkT must be dimensionless, because it is in the exponent. This requires

$$\left[\frac{hc}{\lambda kT}\right] = \frac{\text{J s} \times \text{m/s}}{\text{m} \times \text{J/K} \times \text{K}} = 1 \quad , \tag{4.4}$$

implying that we must input wavelength into the equation in units of meters. The units of hc^2/λ^5 (with λ in m) are

$$\left[\frac{h\,c^2}{\lambda^5}\right] = \frac{\text{J s} \times (\text{m/s})^2}{\text{m}^5} = \frac{\text{W}}{\text{m}^3} \quad . \tag{4.5}$$

The Planck equation is interpreted as a spatial power density per unit wavelength interval. To convert the units of M_λ to W cm^{-2} μm^{-1}, multiply by a factor of 10^{-10} after the calculation is performed. Evalutating M_λ(3 μm, 900 K), we find that hc/λkT = 5.31; (exp{5.31} - 1) = 201.35. Thus, $M_\lambda = 7.6 \times 10^9$ W/m^3 = 0.76 W cm^{-2} μm^{-1}, in good agreement with our initial estimate from the plot.

The Stefan-Boltzmann Law relates the total exitance at all wavelengths to the source temperature, integrating out the wavelength dependence of the blackbody curve:

www.EngineeringEbooksPdf.com

$$M = \int_0^\infty M_\lambda(\lambda,T)\, d\lambda = \int_0^\infty \frac{2\pi h c^2}{\lambda^5 \left(\exp\{hc/\lambda kT\} - 1\right)} d\lambda = \sigma T^4 \quad , \tag{4.6}$$

where σ is the Stefan-Boltzmann constant. The total exitance of a blackbody source increases as the fourth power of the absolute temperature, when the integration is taken over all wavelengths. When the range of integration is limited, the exponent of the dependence of the in-band exitance on temperature can be greater or smaller than 4, because some portions of the spectral-exitance curve change faster than T^4, and some portions change more slowly.

The Wien Displacement Law describes the wavelength of peak exitance of a blackbody as a function of temperature. It can be derived by setting the derivative of the Planck equation with respect to wavelength equal to zero

$$\frac{\partial M_\lambda(\lambda,T)}{\partial \lambda} = 0 \tag{4.7}$$

producing the Wien-Law relationship of peak-exitance wavelength and temperature

$$\lambda_{max} \approx \frac{2898\ [\mu m\ K]}{T\ [K]} \quad . \tag{4.8}$$

For example, a blackbody source at room temperature (300 K) will have a wavelength of peak exitance at around 10 μm. A hotter source, e.g. at 1000 K, will have a shorter peak-exitance wavelength; in this case around 3 μm. Additional insight to the behavior of blackbody sources can be found from the power relationships relative to the half-power points of the Planck function. If we define the half-power point on the short-wavelength side of the peak as λ_{short}, and the corresponding half-power point on the long-wavelength side as λ_{long}, the following relationships can be found

$$\lambda_{short} \approx \frac{1800\ [\mu m\ K]}{T\ [K]} \tag{4.9}$$

$$\lambda_{long} \approx \frac{5100\ [\mu m\ K]}{T\ [K]} \quad . \tag{4.10}$$

For example, a 300-K blackbody has its half-power points at 6 μm and 17 μm. The area under the Planck curve from $\lambda = 0$ to λ_{short}, is 3% of σT^4. The area from λ_{short} to λ_{max} is 25% of σT^4. The area from λ_{max} to λ_{long} is 35% of σT^4, and the area from λ_{long} to ∞ is 37% of σT^4. The Planck function is seen to be very steep at short wavelengths, relatively slowly varying at long wavelengths, and has 60% of its output between the short- and long-wavelength half-power points.

4.3 Emissivity

Although the assumption of blackbody radiation provides an initial model for real sources, a correction factor is often applied to improve the accuracy of flux-transfer calculations. This factor is the emissivity, ε, a dimensionless number that is the ratio of

www.EngineeringEbooksPdf.com

the radiation from the actual source to the radiation from a blackbody at the same temperature. Some definitions of emissivity are based on the various radiation quantities of Table 3.3. A commonly used definition is that of the hemispherical emissivity, written in terms of the ratio of exitance quantities:

$$\varepsilon(\lambda, T) \equiv \frac{M_\lambda(\lambda, T)_{source}}{M_\lambda(\lambda, T)_{blackbody}} \quad . \tag{4.11}$$

In general, ε is a function of both wavelength and temperature, but is typically a slow function of both, so for engineering calculations, the usual specification for emissivity is just a single number for a given material. We see from Eq. (4.11) that emissivity is a number between 0 and 1. A blackbody source has the maximum emissivity of unity.

A graybody source has an emissivity that is less than unity and independent of wavelength. The radiation spectrum of a graybody source is identical to a blackbody source of the same temperature. The peak exitance is at the same wavelength and the spectral exitance at any wavelength is a constant fraction, ε, of what a blackbody source at the same temperature would produce. In Fig. 4.6, we show the spectral exitance as a function of wavelength for a series of 2000-K sources: a blackbody with $\varepsilon = 1.0$, and graybodies with $\varepsilon = 0.75$, 0.5, and 0.25.

Emissivity values for some common materials are seen in Table 4.1. Polished metal surfaces typically have a high reflectivity, a low absorption, and (by Kirchoff's Law) a correspondingly low emissivity. The emissivity is strongly dependent on surface finish, oxidation, or surface contamination. The emissivity values quoted are valid in the long-wave infrared portion of the spectrum, and for temperatures around 300 K. Emissivity tends to increase with increasing temperature, and tends to decrease at longer wavelengths.

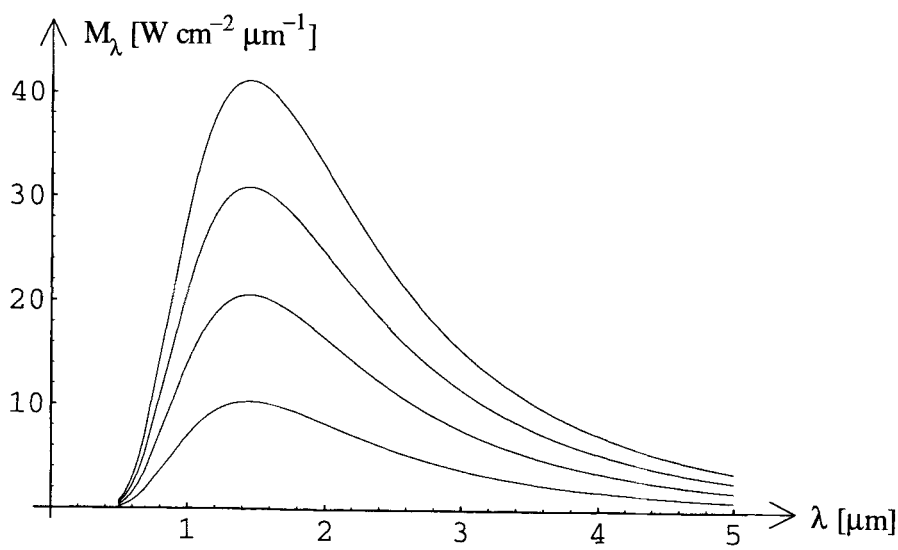


Figure 4.6. $M_\lambda(\lambda, 2000\ K)$ curves for $\varepsilon = 1.0, 0.75, 0.5,$ and 0.25.

www.EngineeringEbooksPdf.com

Table 4.1. Emissivity values for some common materials.

Material	Emissivity
Aluminum	
Polished	0.05
Copper	
Polished	0.05
Oxidized	0.78
Nickel	
Polished	0.05
Oxidized	0.37
with 25 μm oil film	0.27
with 125 μm oil film	0.72
Stainless steel	
Polished	0.07
Oxidized	0.79
Brick	0.93
Concrete	0.92
Sand	0.90
Water	0.96
Human skin	0.98

Bibliography

Dereniak, E., and G. Boreman, *Infrared Detectors and Systems*, Wiley, New York, 1996.
Hudson, R., *Infrared Systems Engineering*, Wiley, New York, 1969.
Taylor, J., *Radiation Exchange*, Academic Press, Orlando, 1990.
Wolfe, W., and G. Zissis, *The Infrared Handbook*, Environmental Research Institute of Michigan, Ann Arbor, 1978.

www.EngineeringEbooksPdf.com

Copyrighted Materials
Copyright © 1998 SPIE Retrieved from www.knovel.com

Chapter 5
Detectors

5.1 Introduction

Detectors are transducers: they produce a measurable electrical output in response to radiation intercepted by the sensing element. The three overall quantities of concern for all detectors are responsivity, response speed, and sensitivity.

Responsivity is a measure of output per unit input. Because there are various input and output quantities, there are a variety of ways to specify responsivity: for example, in volts per watt or in amps per photon per second. Responsivity ($\mathcal{R}$) allows prediction of the magnitude of the sensor's response, given a radiometric calculation of flux on the sensor. Thus, responsivity determines the voltage levels involved at the interface between the detector and the preamplifier that follows.

The response speed of a detector is pertinent because any signal of interest will vary with time. How fast can the signal flux vary and still have the detector follow the variation? The Fourier transform of the time-domain impulse response is the transfer function, the relative response of the sensor as a function of temporal frequency.

Sensitivity is a separate quantity from responsivity. While responsivity is a measure of the output level for a given level of input flux, sensitivity specifies the signal-to-noise ratio (SNR) that the user can expect for a given input flux level. The SNR is a crucial parameter in the determination of image detectability, that is, whether a given feature in the image can be reliably discerned above the noise.

The two primary classes of detectors are thermal detectors and photon detectors. Both kinds of detectors respond to absorbed photons, but their mechanism of response differs, leading to differences in response speed and responsivity as a function of wavelength.

Thermal detectors absorb the energy of the photon as heat. This heat causes a temperature rise in the sensing element. The sensing element has some temperature-dependent electrical property, such as resistance. The change in this electrical property as a function of input flux level is measured by an external circuit.

Photon detectors use the energy of the photon not as heat, but to increase the energy of a charge carrier, so that the carrier makes an electronic transition across a forbidden energy gap. This is typically a transition of an electron from the valence band to the conduction band in a semiconductor material. The excitation of these carriers into a higher energy state affects the sensor's electrical properties. The change of electrical properties as a function of input flux level is measured by external circuitry.

We compare the properties of thermal and photon sensors in terms of response speed and spectral responsivity $\mathcal{R}(\lambda)$. Thermal detectors are slow because a finite time is required for the sensing element to rise in temperature after the absorption of energy. Typical time constants for thermal detectors are in the range of tenths of seconds to milliseconds. A trade-off exists between response speed and response magnitude for thermal detectors. Because of their relatively long thermal time constants, waiting a longer time generally produces a larger response from a thermal detector. Photon detectors are fast, because an electronic transition is virtually instantaneous upon photon absorption. Typical time constants for photon detectors are in the microsecond range and shorter. Response speed is often determined by the resistance-capacitance (RC) product of the external readout circuit that interfaces to the photon detector.

www.EngineeringEbooksPdf.com

5.2 Cutoff Wavelength

In terms of spectral responsivity $\mathcal{R}(\lambda)$, responsivity as a function of wavelength of the incoming radiation, a thermal detector has a flat response, when plotted in energy-based units. As seen in Fig. 5.1, the ideal thermal sensor has a responsivity that is independent of wavelength because a watt of radiation absorbed by the sensor at $\lambda = 1$ μm will cause the same temperature rise in the sensing element as a watt of radiation at $\lambda =$ 10 μm. There is no long-wavelength cutoff behavior in thermal sensors because no energy gap is inherent in the mechanism. Practically, the wavelength range for a thermal sensor is limited by the spectral absorption of the sensor material and the spectral transmission range of the window material that typically covers the sensor.

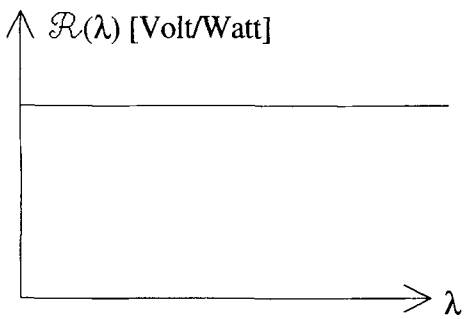


Figure 5.1. Spectral responsivity for a thermal detector.

A photon detector has an inherent nonuniformity of response as a function of wavelength. For a photon incident on the sensor to be absorbed by the material and impart its energy to an electron, its energy must be sufficient to lift the electron across the energy gap. Recall that Eq. (3.4)

$$\mathcal{E} = hc/\lambda \tag{5.1}$$

implies that the photon energy $\mathcal{E}$ [Joules/photon] is inverse with wavelength. As seen in Fig. 5.2, given a photon sensor with an energy gap $\mathcal{E}_{gap}$, photons with wavelength longer than the long-wavelength cutoff $\lambda_{cut} = hc/\mathcal{E}_{gap}$ are not absorbed and not detected.

The wavelength where the photon has just enough energy to bridge $\mathcal{E}_{gap}$ corresponds to the long-wavelength cutoff $\lambda_{cut} = hc/\mathcal{E}_{gap}$, and is the longest wavelength that will be detected by a sensor with a given $\mathcal{E}_{gap}$. We can solve for the long-wavelength cutoff of the photon detector in terms of $\mathcal{E}_{gap}$:

$$\mathcal{E}_{gap}[\text{eV}] = \frac{hc}{\lambda_{cut}} = \frac{6.6 \times 10^{-34}\ \text{J s}\ \ 3 \times 10^{8}\ \text{m/s}}{\lambda_{cut}\ \mu\text{m} \times 10^{-6}\ \text{m/}\mu\text{m}} \times \frac{1\ \text{eV}}{1.6 \times 10^{-19}\ \text{J}} = \frac{1.24}{\lambda_{cut}\ [\mu\text{m}]} \tag{5.2}$$

www.EngineeringEbooksPdf.com

If Si, with $\mathscr{E}_{gap}$ = 1.12 eV, is used as a sensor, photons with wavelengths shorter than 1.1 μm will be detected. If InSb, which has a smaller $\mathscr{E}_{gap}$ = 0.22 eV, is used, the sensor will detect photons with wavelengths shorter than 5.6 μm.

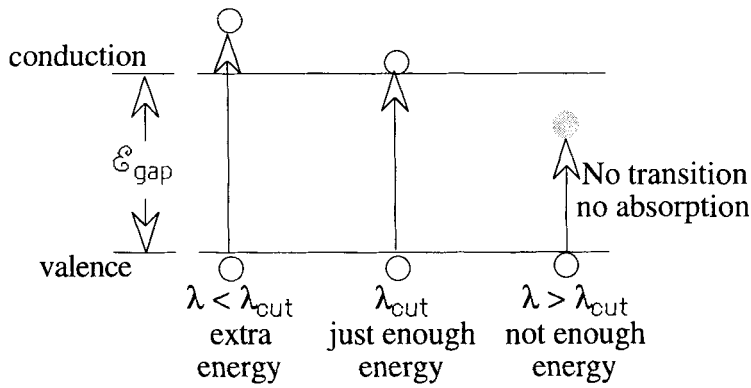


Figure 5.2. Origin of cutoff wavelength for a photon detector.

5.3 Cooling Requirements

To detect long-wavelength photons, the sensor element in photon detectors must often be cooled. Charge carriers in a semiconductor material have an energy that is proportional to kT, where k is Boltzmann's constant (8.62×10^{-5} eV/K). In a photon detector, the thermal energy of the charge carriers can cause excitation of electrons across the energy gap of the sensor without the detection of photons, a phenomenon known as dark current. Once the electrons have been raised in energy, thermally excited electrons are indistinguishable from photogenerated electrons. There is a shot noise associated with dark current. We will see later in this chapter that the shot-noise current has an rms value proportional to the square root of the dark current.

The sensor element must be kept at a sufficiently low temperature that the charge carriers have a statistically small probability of having enough thermal energy to cross the energy gap. The detection of long-wavelength photons requires smaller energy gaps. Thus, lower temperatures are needed for supression of the dark current. In Table 5.1 we compare kT to $\mathscr{E}_{gap}$ for various photon detectors.

Table 5.1. Comparison of kT to energy gap for photon sensors.

Si:	λ_{cut}= 1.1 μm	$\mathscr{E}_{gap}$ = 1.12 eV	kT @ 300 K = 0.026 eV = 2.3% of $\mathscr{E}_{gap}$
InSb:	λ_{cut}= 5.6 μm	$\mathscr{E}_{gap}$ = 0.22 eV	kT @ 180 K = 0.015 eV = 6.8% of $\mathscr{E}_{gap}$ kT @ 300 K = 11.8% of $\mathscr{E}_{gap}$
HgCdTe	$\lambda_{cut} \approx$ 10 μm	$\mathscr{E}_{gap} \approx$ 0.12 eV	kT @ 77 K = 0.0066 eV = 5.5% of $\mathscr{E}_{gap}$ kT @ 300 K: 22% of $\mathscr{E}_{gap}$

www.EngineeringEbooksPdf.com

While Si can be operated at room temperature, neither InSb nor HgCdTe has acceptable levels of dark current at 300 K. InSb requires at least 180 K, a temperature typically reached with thermoelectric coolers, while HgCdTe requires cooling to 77 K, typically obtained with open-cycle liquid nitrogen or mechanical refrigerators.

5.4 Spectral Responsivity

As seen in Fig. 5.1, the spectral responsivity curve for a thermal detector is plotted in energy-derived units (e.g., V/W or A/W). Ideally, these curves are constant with wavelength, and do not exhibit a long-wavelength cutoff. Spectral responsivity for ideal photon detectors goes to zero for wavelengths longer than the cutoff wavelength of Eq. (5.2), but the curve can be plotted with respect to either energy-derived or photon-derived units. Photon detectors are most naturally described in photon-derived units but for historical reasons are often plotted with respect to energy-derived units. Because the proportionality between these two sets of units, $\mathscr{E} = hc/\lambda$, depends on wavelength, these two sets of units will produce responsivity curves of different shape, as seen in Fig. 5.3.

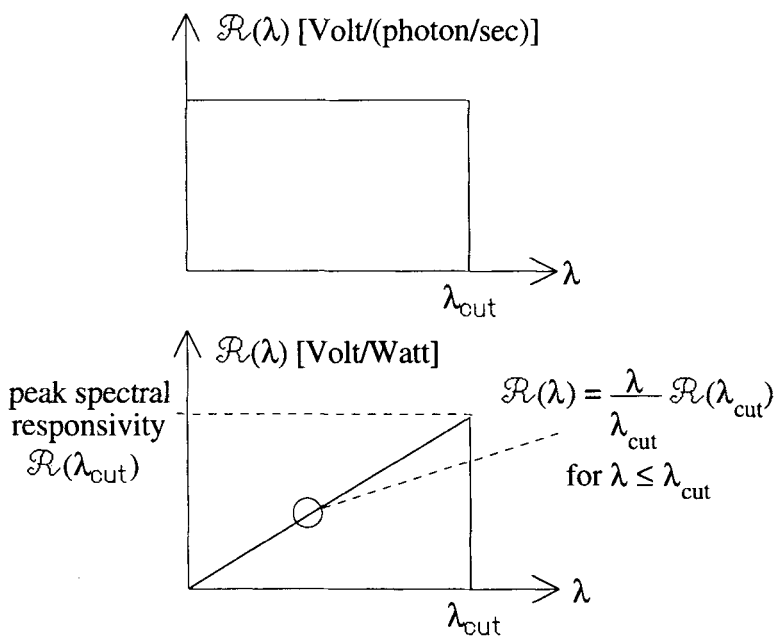


Figure 5.3. Spectral responsivity for an ideal photon detector, plotted in photon units and in energy units.

An ideal photon detector has a flat $\mathscr{R}(\lambda)$ out to λ_{cut} when plotted in photon-derived units (e.g., A/(photon/s). Each photon produces the same amount of response, as long as the photon has enough energy to bridge the gap. If $\mathscr{R}(\lambda)$ is plotted as a function of energy-derived units, a linear increase in responsivity is seen up to λ_{cut}. It takes more photons at long wavelengths to make 1 W of power and the detector responds the same to each photon up to λ_{cut}. For photons with $\lambda = 1$ μm, 1 W = 5×10^{18} photons/s, for

www.EngineeringEbooksPdf.com

photons with $\lambda = 10\ \mu m$, $1\ W = 50 \times 10^{18}$ photons/s. If the sensor responds equally on a per-photon basis, it will appear to have a factor of 10 higher responsivity at 10 μm than at 1 μm, if the spectral responsivity is plotted in energy-derived units. In this case, an ideal photon detector has a spectral responsivity of the form

$$\mathcal{R}(\lambda) = \frac{\lambda}{\lambda_{cut}} \mathcal{R}(\lambda_{cut}) \text{ for } \lambda < \lambda_{cut} \; ; \; \mathcal{R}(\lambda) = 0 \text{ for } \lambda > \lambda_{cut} \quad . \tag{5.3}$$

The significance of the spectral responsivity from a design viewpoint is that it is used to calculate the output from a detector in response to flux from a spectrally distributed source

$$\text{Output} = \int_0^\infty \phi_\lambda(\lambda)\, \mathcal{R}(\lambda)\, d\lambda$$
$$[V] = [W/\mu m]\,[V/W]\,[\mu m] \quad . \tag{5.4}$$

In the overlap integral of Eq. (5.4), there is a contribution to detector output only at those wavelengths where both $\phi_\lambda(\lambda)$ and $\mathcal{R}(\lambda)$ are nonzero. The spectral flux falling on the detector can be calculated from spectral exitance as

$$\phi_\lambda(\lambda) = \frac{M_\lambda(\lambda)}{\pi} A_s \Omega_d \quad . \tag{5.5}$$

As an example of calculation of the output of an ideal photon detector and a blackbody source, when both are specified in energy-based units, Eq. (5.4) becomes

$$\text{Output} = \frac{A_s \Omega_d}{\pi} \int_0^{\lambda_{cut}} \frac{2\pi h c^2}{\lambda^5 \left(\exp\{hc/\lambda kT\} - 1\right)} \frac{\lambda}{\lambda_{cut}} \mathcal{R}(\lambda_{cut})\, d\lambda \quad . \tag{5.6}$$

For a thermal detector (flat $\mathcal{R}$ as a function of λ) and a blackbody source, when both are specified in energy-based units, Eq. (5.4) becomes

$$\text{Output} = \frac{A_s \Omega_d}{\pi} \int_0^\infty \frac{2\pi h c^2}{\lambda^5 \left(\exp\{hc/\lambda kT\} - 1\right)} \mathcal{R}\, d\lambda = \frac{A_s \Omega_d}{\pi} \sigma T^4 \mathcal{R} \quad . \tag{5.7}$$

Response of a detector to a laser source can also be calculated using Eq. (5.4), by taking the spectral flux from the laser to be approximately a delta function

$$\phi_\lambda(\lambda) = \phi_{laser}\, \delta(\lambda - \lambda_{laser}) \quad . \tag{5.8}$$

Using the sifting property of the delta function, the detector output is then

www.EngineeringEbooksPdf.com

$$\text{Output} = \int_0^\infty \phi_{\text{laser}}\, \delta(\lambda - \lambda_{\text{laser}})\, \mathcal{R}(\lambda)\, d\lambda = \phi_{\text{laser}} \mathcal{R}(\lambda_{\text{laser}}) \quad . \tag{5.9}$$

5.5 Frequency Response and Noise-Equivalent Bandwidth

A frequency-domain transfer function can be obtained for a sensor by the Fourier transform of the time-domain impulse response. This transfer function will multiply the spectral responsivity developed in Section 5.4. Two particular forms of the impulse response are of interest for sensor analysis. The first is the decaying-exponential impulse response seen in Eq. (5.10):

$$v(t) = v_0\, e^{-(t-t_0)/\tau} \quad . \tag{5.10}$$

This type of impulse response naturally arises from charge-carrier-lifetime effects in semiconductors and from the RC time constant of electronic circuits, where τ is interpreted as the carrier lifetime or the RC product. Taking the magnitude of the Fourier transform of the impulse reponse in Eq. (5.10), we obtain the voltage transfer function

$$\mathcal{R}(f) = \frac{1}{\sqrt{1 + (2\pi\tau f)^2}} \quad , \tag{5.11}$$

which is plotted in Fig. 5.4.

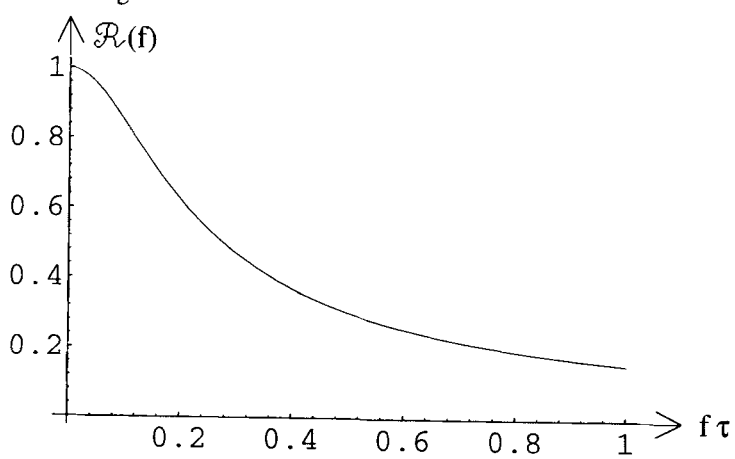


Figure 5.4. Frequency response for decaying-exponential impulse reponse.

The integrator is a second type of impulse response important for detectors. The impulse response seen in Eq. (5.12) corresponds to a system that integrates a signal for a time period τ, producing a square-pulse response

$$v(t) = v_0\, \text{rect}\left(\frac{t - t_0}{\tau}\right) \quad . \tag{5.12}$$

Calculation of the transfer function for this impulse response produces the sinc-function form seen in Eq. (5.13), which is plotted in Fig. 5.5

$$\mathcal{R}(f) = \left|\frac{\sin(\pi\tau f)}{\pi\tau f}\right| \quad . \tag{5.13}$$

www.EngineeringEbooksPdf.com

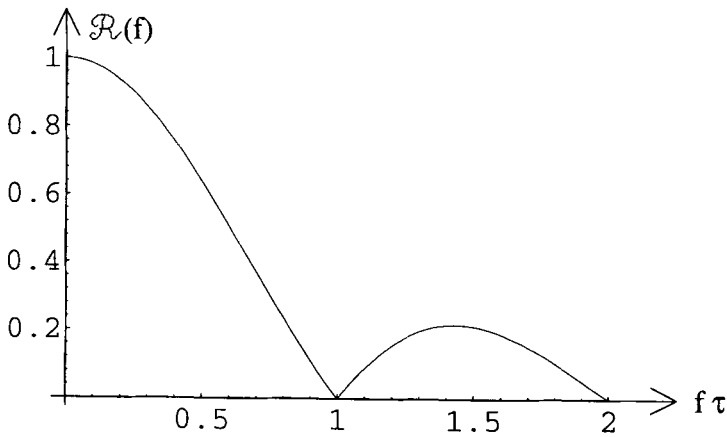


Figure 5.5. Frequency response for integrator impulse reponse.

As an example of the use of the frequency response, suppose we have a sensor with a dc responsivity of 5 mV/mW, and an exponential time response with a time constant τ = 133 μs. If a steady-state flux of 10 mW falls on the detector, its output is 50 mV. If the incoming flux is modulated at f = 25 Hz, the transfer function of Eq. (5.11) will be approximately 1, and the modulated peak-to-peak signal amplitude is still 50 mV. If the flux is modulated at 1200 Hz, then $\mathcal{R}(f) = 0.707$, and the signal amplitude is decreased to 33.35 mV. If the flux is modulated at 1800 Hz, the signal level is further decreased to 27.75 mV.

The frequency response is also used to calculate the noise-equivalent bandwidth (Δf) of a system. The noise-equivalent bandwidth is the width of a flat-bandpass filter, seen in Fig. 5.6, that will pass the same amount of white-noise power as the original transfer function.

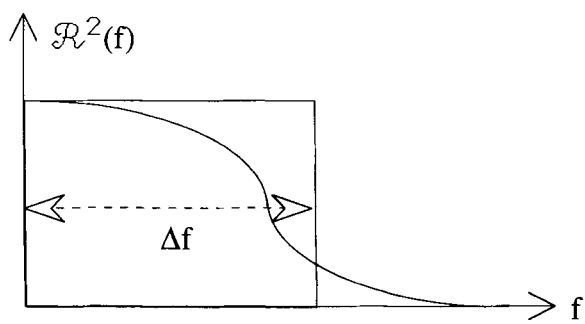


Figure 5.6. Noise-equivalent bandwidth.

The calculation of the noise-equivalent bandwidth uses the square of the original voltage transfer function, because power is proportional to the voltage squared

$$\Delta f = \int_0^\infty \left| \frac{\mathcal{R}(f)}{\mathcal{R}(f=0)} \right|^2 df \qquad (5.14)$$

www.EngineeringEbooksPdf.com

Carrying out the integration in Eq. (5.14) for the exponential-impulse-response transfer function of Eq. (5.11) yields

$$\Delta f = 1/(4\tau) \tag{5.15}$$

and for the integrator transfer function of Eq. (5.13), the noise-equivalent bandwidth is

$$\Delta f = 1/(2\tau) \ . \tag{5.16}$$

Equation (5.16) is often used in noise analyses, assuming the integrator form of the transfer function. Both Eqs. (5.15) and (5.16) imply that a brief impulse response implies a broad bandwidth. Conversely, more time must be taken to make a measurement with a narrow bandwidth than one with a wide bandwidth.

5.6 Noise Terminology

There is always some fluctuation in any electrical quantity being measured. The small sizes of the typical voltages and currents involved in detection of optical or infrared radiation dictate that we consider sources of noise. Noise may arise in the sensor itself, in the amplifier, or as external interference. In high-performance systems, the detector (or detector array) is often a component whose cost strongly influences the overall cost of the system. In such a situation, the amplifier is usually designed so that the sensitivity of the system is detector-noise limited. For lower-cost systems that have reduced sensitivity requirements, the system sensitivity will often be limited by noise in the amplifier.

We can describe the time-domain noise in terms of either voltage or current. If, for example, we designate a random-noise voltage waveform $v_n(t)$ [volts], three main descriptors exist for this noise waveform: (1) the mean value

$$\overline{v_n}\ [\text{volts}] = \frac{1}{T}\int_0^T v_n(t)\,dt \ , \tag{5.17}$$

(2) the mean-square (ms) value, in units of volts^2 or power; the variance of the waveform

$$\sigma_n^2\ [\text{volts}^2] = \frac{1}{T}\int_0^T \left(v_n(t) - \overline{v_n}\right)^2 dt \ , \tag{5.18}$$

and (3) the root-mean-square (rms) value; the standard deviation of the waveform

$$\sigma_n\ [\text{volts}] = \sqrt{\frac{1}{T}\int_0^T \left(v_n(t) - \overline{v_n}\right)^2 dt} \ . \tag{5.19}$$

Noise contributions arising from separate (uncorrelated) sources must be added in terms of power. This so-called addition in quadrature adds the mean-square (power) quantities and then takes the square root of the sum to yield an rms value

$$\text{rms}_{\text{total}} = \sqrt{(\text{rms}_1)^2 + (\text{rms}_2)^2} \neq (\text{rms}_1)^2 + (\text{rms}_2)^2 \ . \tag{5.20}$$

For example, if two independent voltage noise sources of 3 V rms each are added in series, as seen in Fig. 5.7, the result is 4.24 V rms.

www.EngineeringEbooksPdf.com

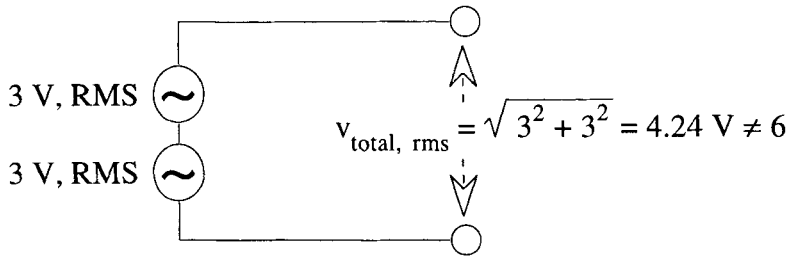


Figure 5.7. Addition of two uncorrelated noise sources in quadrature.

For optical and infrared sensors, we calculate a signal-to-noise ratio (SNR) as a peak signal voltage (or current) divided by an rms noise voltage (or current). The rms noise voltage (or current) is the addition of all the significant noise sources in quadrature. In optics, the usual SNR is a voltage (or current) SNR rather than a power SNR, which is common in the specification of radio-frequency systems.

Noise can also be decribed in terms of a noise power spectrum [W/Hz], a power per unit frequency, also called the power spectral density (PSD). The PSD, seen in Fig. 5.8, gives insight into the origin of the various noise components because of their differing frequency contents. This frequency dependence of the noise affects the choice of modulation frequency for a narrowband (e.g., communication) system or affects the choice of the lower and upper bandwidth limits for a wideband (e.g., electronic-imaging) system. At low frequencies (less than 1 kHz), 1/f noise generally dominates. In the midband range of frequencies the main noise contribution is either shot noise or generation-recombination (G-R) noise. The PSD in this midrange is generally flat from the 1/f rolloff frequency out to the inverse of the carrier lifetime at about 20 kHz to 1 MHz. Beyond that carrier-lifetime rolloff frequency, the main noise contributions are Johnson noise and amplifier noise. These are generally flat, even to the highest frequencies of interest for sensor systems. The passband for electronic-imaging systems is typically set to utilize the flat midband region of the spectrum. The charge-carrier rolloff frequency is an upper limit for the signal, because the signal is necessarily transmitted by charge carriers. A system bandwidth with a higher cutoff frequency will pass excess levels of noise with no additional signal. The lower limit of the system passband typically rolls off the 1/f noise, making an ac-coupled system with little or no response at dc.

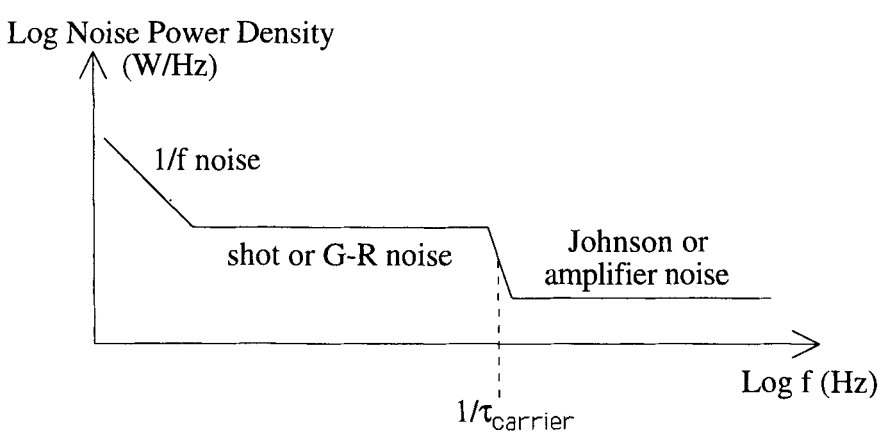


Figure 5.8. Plot of power spectral density provides insight into noise mechanisms.

www.EngineeringEbooksPdf.com

The assumption of white noise (flat PSD) is often made in engineering calculations. White noise power scales with bandwidth, as seen in Fig. 5.9, and thus rms noise scales as the square root of the bandwidth. For example, if a particular noise source has a noise PSD of 50 nW/Hz, then in a 1-Hz bandwidth, the source has 50×10^{-9} W of noise power. Over a Δf of 10 Hz, the source yields 500 nW, and over a 500 KHz bandwidth there is 25 mW of noise power.

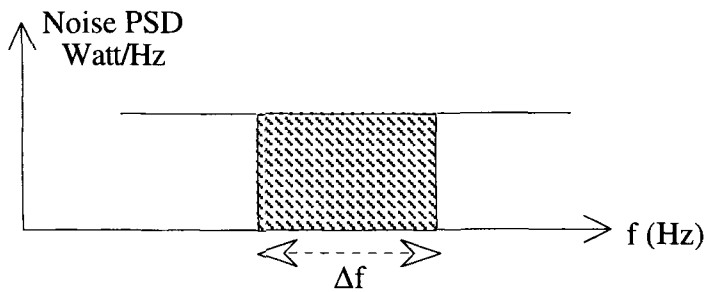


Figure 5.9. White-noise power is proportional to bandwidth.

Oftentimes the spectrum of the noise is discussed in terms of rms quantities, as a noise-voltage spectrum ($V/Hz^{1/2}$) or a noise-current spectrum ($A/Hz^{1/2}$). As noted above, these rms quantities scale with the square root of the bandwidth. For example, if a noise source has an rms noise of 7.5 μV in a 50-Hz bandwidth, the noise-voltage spectrum is $7.5 \times 10^{-6}/50^{1/2} = 1.1 \times 10^{-6}$ $V/Hz^{1/2}$. Over a bandwidth of 500 kHz, this source will yield an rms noise voltage of $1.1 \times 10^{-6} \times 500{,}000^{1/2} = 7.8 \times 10^{-4}$ V.

5.7 Shot Noise and Generation-Recombination Noise

Shot noise arises because both photons and charge carriers are quantized. Neither an electron nor a photon can be divided. In the context of photon detection, a photon either excites an electron or it does not. This quantization is the origin of a fluctuation that affects the SNR, noticed primarily at low levels of photon or electron flow. Shot noise is present in devices such as photovoltaics or vacuum photodiodes, where the photogenerated charge carriers must cross a potential-energy barrier in order to be detected. Carrier generation is a random process, depending on the statistics of photon arrivals and the quantum efficiency. However, in these diode-like devices, carrier recombination is a nearly deterministic process, with the potential-energy-barrier transit time determining an effective carrier lifetime. Shot noise is replaced by generation-recombination noise in devices such as photoconductors that lack the potential-energy barrier. Shot noise is zero in devices that have a zero current flow, such as an open-circuit photovoltaic detector.

The rms shot-noise current σ_i (amps) that arises from a dc current flow of i_{avg} is:

$$\sigma_i = \sqrt{2\, q\, i_{avg}\, \Delta f}\ , \tag{5.21}$$

where q is the charge on an electron equal to 1.6×10^{-19} Coulomb, and Δf is the noise-equivalent bandwidth of the noise measurement, generally taken as $1/(2\tau)$, where τ is the measurement interval. Electrons are particles that obey Poisson statistics, hence the square-root dependence of the standard deviation on the mean current. For an average

www.EngineeringEbooksPdf.com

current flow of 1 mA, the rms noise current is 2×10^{-10} A in a 100-Hz bandwidth. Let us examine the two functional dependences for shot noise: bandwidth and mean current level. The square-root-of-bandwidth dependence can be interpreted to mean that more averaging (longer measurement interval, narrower noise bandwidth) results in less noise, as seen in Fig. 5.10.

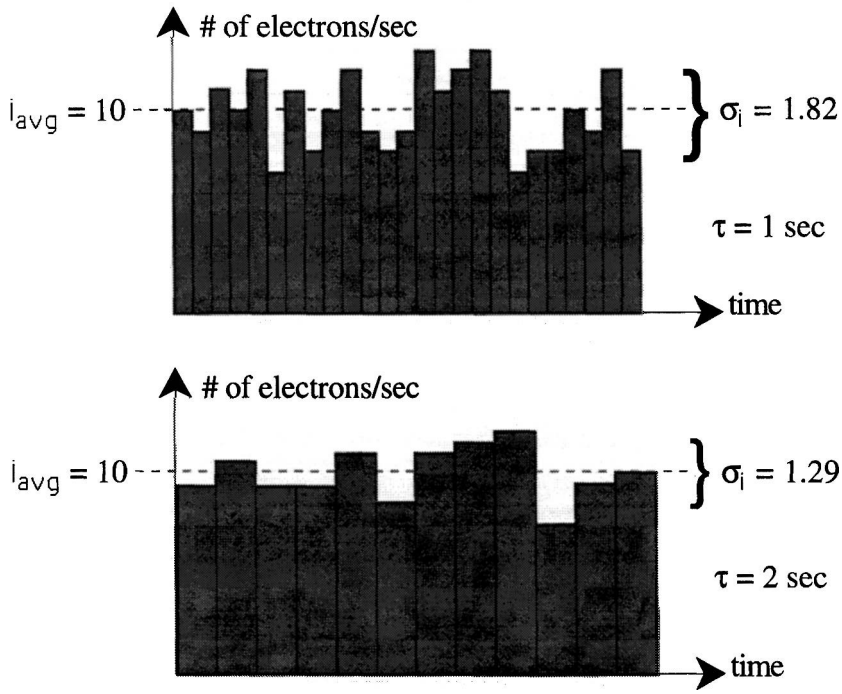


Figure 5.10. Longer measurement times mean less shot noise.

The other square-root dependence is that of rms noise on the average current level. A larger dc current implies a larger rms variation, as seen in Fig. 5.11, but shot-noise fluctuations are noticed primarily at low current levels. Lower levels of current have smaller rms noise, but also are seen to have a lower SNR.

Assume that the average current flowing is entirely caused by the signal that we want to detect. Then the shot noise on the signal current is the dominant noise source. This situation is called the signal-shot-noise limit. We see in Fig. 5.11 that when the mean current level is increased, the rms noise level increases as the square root of the mean. Larger current means more noise. However, the signal-shot-noise-limited SNR is

$$\mathrm{SNR}_{\text{signal-shot-limit}} = \frac{i_{sig}}{i_{\text{noise, rms}}} = \frac{i_{sig}}{\sqrt{2\, q\, i_{sig}\, \Delta f}} \propto \sqrt{i_{sig}} \quad . \tag{5.22}$$

Equation (5.22) shows that the shot noise is most troublesome (yielding the lowest value of SNR) at low current levels. Even though the numerical value of the rms noise grows

www.EngineeringEbooksPdf.com

with increasing current, the SNR increases because while noise grows as the square root of the signal current, the signal grows linearly, producing an overall square-root dependence for SNR.

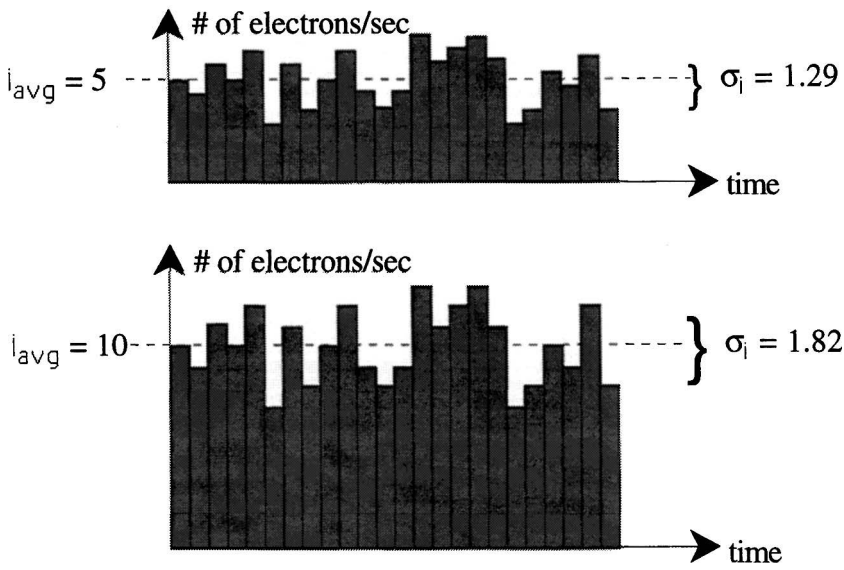


Figure 5.11. Larger average current means more shot noise.

In addition to the generation of electrons by signal photons, there are other electron-generation processes that often contribute to the total current, as seen in Fig. 5.12. The total current can include dark-current electrons (generated by thermal excitation of carriers over the energy gap of the detector) and electrons generated by detection of background photons

$$i_{total} = i_{sig} + i_{bkg} + i_{dark} \quad . \tag{5.23}$$

The dc portion of the nonsignal contributions can be subtracted from the total current. However, each source of excess carriers beyond those generated from signal photons produces additional current fluctuation which does not subtract as a dc level. Shot noise is associated with each of these excess-current sources, reducing the overall SNR.

For the general case of shot-noise-limited performance we find

$$SNR_{\text{shot-noise-limit}} = \frac{i_{sig}}{\sqrt{2\,q\,i_{total}\,\Delta f}} = \frac{i_{sig}}{\sqrt{2\,q\left(i_{sig} + i_{bkg} + i_{dark}\right)\Delta f}} \quad . \tag{5.24}$$

Making Eq. (5.24) specific to the case of background-shot-noise-limited detection, we find

www.EngineeringEbooksPdf.com

$$SNR_{bkg\text{-}shot\text{-}noise\text{-}limit} = \frac{i_{sig}}{\sqrt{2\, q\, i_{bkg}\, \Delta f}}\ , \tag{5.25}$$

and for the dark-current shot-noise limit, we find

$$SNR_{dark\text{-}shot\text{-}noise\text{-}limit} = \frac{i_{sig}}{\sqrt{2\, q\, i_{dark}\, \Delta f}}\ . \tag{5.26}$$

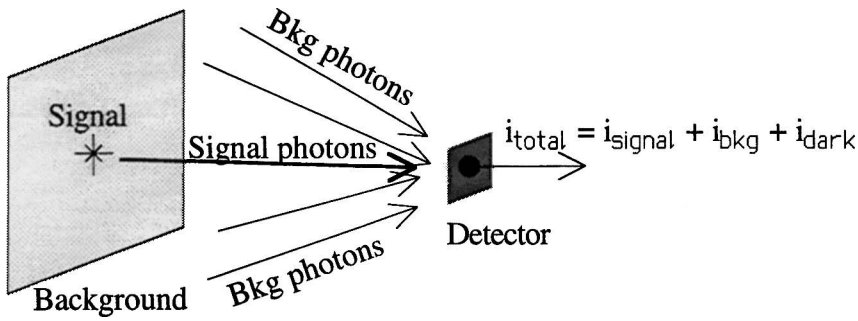


Figure 5.12. Background and dark-current carriers contribute to shot niose.

Equation (5.24) applies only to the case of shot-noise-limited detection. Any other sources of noise, such as amplifier noise or Johnson noise, will add in quadrature in a calculation of the total rms noise.

In detectors such as photoconductors that do not have a potential energy barrier, both generation and recombination are random processes. Thus generation-recombination (G-R) noise has a larger fluctuation than shot noise, because it is the summation of two random processes. The two processes add their noises in quadrature, producing an rms noise which is the square root of two larger than that indicated by the shot noise expression, Eq. (5.21). When all other factors (e.g., photon flux level, quantum efficiency, bandwidth) are held constant, photovoltaic sensors have a higher SNR than do photoconductors.

5.8 Johnson Noise and 1/f Noise

Charge carriers in a resistive element have thermal energy at any temperature above absolute zero. This thermal energy is manifest as Brownian motion of the carriers. Higher temperatures produce more vigorous motion. Across the open circuit of the terminals of a resistor, a zero-mean fluctuating voltage is seen with an rms value of

$$v_{Johnson,\ rms} = \sqrt{4\, k\, T\, R\, \Delta f}\ , \tag{5.27}$$

where R is the resistance of the element, T is the temperature in Kelvin, and k is Boltzmann's constant. The (Thevenin) circuit model used is that of a noise-free resistance in series with an rms voltage source described by Eq. (5.27), as seen in Fig. 5.13.

www.EngineeringEbooksPdf.com

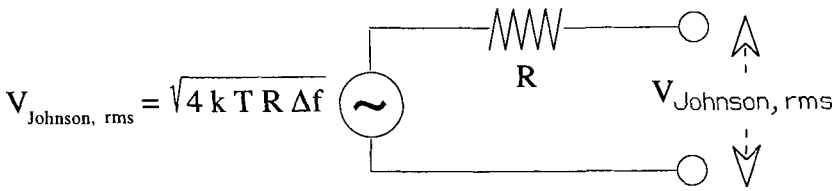


Figure 5.13. Johnson-noise voltage model.

An alternate (Norton) model for the resistor can be used to model the rms noise current that flows when the terminals of the resistor are shorted. An rms current source is taken to be in parallel with a noise-free resistor, where the rms Johnson-noise current is

$$i_{Johnson,\ rms} = \sqrt{\frac{4\,k\,T\,\Delta f}{R}}\ , \tag{5.28}$$

as seen in Fig. 5.14.

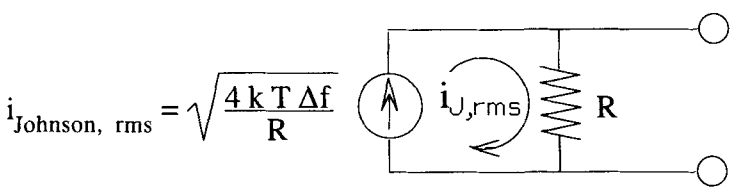


Figure 5.14. Johnson-noise current model.

To show that the models of Figs. 5.13 and 5.14 are two-terminal equivalent, a load resistor is placed across the output terminals, and we find that the same power is dissipated in the load in each case

$$P_{dissip.\ in\ R_L} = \frac{v_{rms}^2}{R_L} = i_{rms}^2\,R_L = \frac{4\,k\,T\,R\,\Delta f}{(R + R_L)^2}\,R_L\ . \tag{5.29}$$

As a numerical example, consider a 50-Ω resistor at room temperature (300 K), where we find that the rms Johnson-noise voltage per root Hz is $(4\ k\ T\ R)^{1/2} = 0.9$ $nV/Hz^{1/2}$. Over a bandwidth of 500 kHz, this resistor yields an rms noise voltage of 0.63 μV. A 1-MΩ resistor at room temperature produces an rms Johnson-noise voltage per root Hz of $(4\ k\ T\ R)^{1/2} = 0.13$ $\mu V/Hz^{1/2}$. Over $\Delta f = 500$ kHz, this resistor yields an rms noise voltage of 0.13 $\mu V/Hz^{1/2} \times (500{,}000)^{1/2} = 92$ μV. For the same resistor, the corresponding Johnson-noise current spectrum at 300 K is 1.3×10^{-13} $A/Hz^{1/2}$.

When two resistors (typically corresponding to a detector resistance and a load resistance) are at different temperatures, the total rms noise is found by quadrature addition. Often the detector is at 77 K while the load (imput impedance of the

www.EngineeringEbooksPdf.com

preamplifier) is at room temperature. Suppose we have two resistors in parallel, a resistance R_1 at a temperature T_1 and a resistance R_2 at a temperature T_2. As seen in Fig. 5.15, we take the Norton model of each resistor, add the parallel current noises in quadrature, and combine the noise-free resistors in parallel as usual.

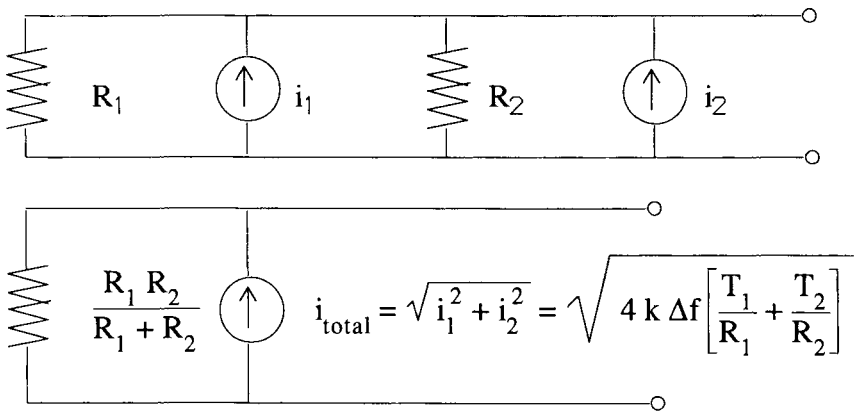


Figure 5.15. Parallel combination of resistors at two different temperatures.

One-over-f noise is present when a bias current flows (such as in bolometers and photoconductors), and is absent in the open-circuit voltage-generation mode of photovoltaic detectors. One-over-f noise has a power spectrum that is inverse with frequency

$$i_n^2 \left[\frac{\text{Amp}^2}{\text{Hz}}\right] \propto \frac{i_{dc}^2 \, \Delta f}{f} \tag{5.30}$$

and is often the dominant noise source at low freqencies, below approximately 1 kHz. The fluctuations are often noticed as a low-frequency drift in the current level. The causes of 1/f noise include the non-ohmic nature of electrical contacts between metals and semiconductors. The usual method for reduction of 1/f noise is the application of a low-frequency electronics filter, which ac couples the current information, rolling off the response at dc.

5.9 Noise Specification of Detectors: Noise-Equivalent Power

We desire a means to specify the noise performance of a detector, so that we can predict the SNR which will be obtained when a given amount of power falls on the detector. Considering Fig. 5.16, the responsivity relates the output of a detector to the incident input flux. If the input flux is decreased, the output level will decrease to a point where the rms noise provides the baseline response, which is independent of flux level.

The noise-equivalent power (NEP) is the amount of flux [W] that would produce an output equal to the rms value of the noise, assuming a linear responsivity curve. Thus, NEP is the input flux required to produce SNR = 1. Noise-equivalent power is thus interpreted as the "minimum detectable power," although SNR = 1 is only a

www.EngineeringEbooksPdf.com

reference level, and the actual input flux required for acceptable sensor operation will depend on the SNR requirements which are usually determined by the false-alarm rate and the probability of detection. The definition of NEP is

$$\text{NEP} = \frac{\phi_{\text{det}}}{\text{SNR}} , \tag{5.31}$$

which shows that NEP is the ϕ_{det} for SNR =1. Equation (5.31) also allows calculation of the required ϕ_{det} for a given system as the product of SNR and NEP. Sensitivity is better for smaller NEP, in that the SNR produced is higher for a given sensor flux level. With reference to the triangle in Fig. 5.16, NEP can also be written in terms of the rms noise and the responsivity as

$$\text{NEP} = \frac{v_{n,\text{rms}}}{\mathcal{R}_v} = \frac{i_{n,\text{rms}}}{\mathcal{R}_i} , \tag{5.32}$$

where $\mathcal{R}_v$ is the voltage responsivity [V/W] and $\mathcal{R}_i$ is the current responsivity [A/W].

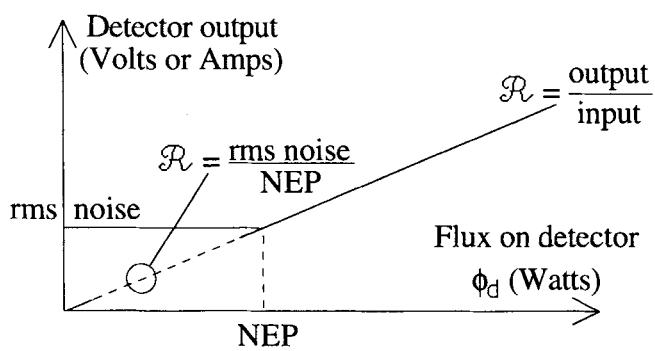


Figure 5.16. Noise-equivalent power and responsivity.

We now consider some numerical-calculation examples with noise-equivalent power. A detector has been measured to have a responsivity of 10 V/W, and an rms noise voltage of 7.5 μV in a bandwidth of 500 kHz. The NEP is calculated using Eq. (5.32) as NEP = [7.5 μV]/[10 V/W] = 7.5×10^{-7} W. Using this NEP and Eq. (5.31), we predict the SNR when 1 μW of signal power falls on the detector as SNR = 1 μW/0.75 μW = 6.7. If the bandwidth is changed from 500 kHz to 20 Hz, both the rms noise and the NEP will change by a factor of the square root of the ratio of the bandwidths. Thus the new rms noise is 7.5 μV × $(20/500{,}000)^{1/2}$ = 47 nV. Assuming that the responsivity is independent of noise bandwidth, the new NEP can be calculated by Eq. (5.32) as NEP = [47 nV]/[10 V/W] = 4.7 nW. The system sensitivity was improved by reduction of the noise bandwidth.

www.EngineeringEbooksPdf.com

5.10 Normalized Detectivity D*

The normalized detectivity, D*, is a figure of merit often used to specify the performance of a detector. D* is inversely proportional to the NEP, so that bigger D* corresponds to better sensitivity. D* is also proportional to the square root of the detector area and the square root of the measurement bandwidth

$$D^* = \frac{\sqrt{A_d}\sqrt{\Delta f}}{NEP} \left[\frac{cm\sqrt{Hz}}{Watt}\right] . \tag{5.33}$$

Normalization with respect to detector area and measurement bandwidth cancels out the dependence of NEP on these quantities. We have seen in Eq. (5.32) that NEP is proportional to the rms noise and hence (for white noise) to the square root of the measurement bandwith. This dependence will cancel in the calculation of D*, yielding a figure of merit independent of the bandwidth used to make the measurement. Similar reasoning is used for the root-area dependence. It is found empirically that the rms noise level measured in a detector is proportional to the square root of the area of the detector. More detector area and hence more volume of detector material tends to generate more noise power. The normalization used in Eq. (5.33) yields a figure of merit independent of both the measurement bandwidth and the detector area. Thus, D* is most useful in the comparison of the merits of different detector materials and fabrication processes, without consideration of a particular bandwidth or area inherent in a given application. Manufacturers of detectors often specify the performance of their generic detector products using D*. Given the D*, an end user can calculate the SNR using a combination of Eqs. (5.33) and (5.31)

$$SNR = \frac{\Phi_{det}}{NEP} = \frac{\Phi_{det}\ D^*}{\sqrt{A_d}\sqrt{\Delta f}} . \tag{5.34}$$

The D* is a consideration when choosing an appropriate detector technology for a particular application. However, the NEP remains the fundamental quantity for the calculation of SNR. Calculation of SNR requires a specification of the detector area for a particular application, in consideration of requirements of field of view and the detector footprint as detailed in Section 1.6. The required update rate of the sensor system will ultimately determine the noise bandwidth. Although their effect on D* has been normalized out according to Eq. (5.33), both the detector area and the bandwidth affect the final SNR achieved by the sensor system.

Other useful relationships involving D* can be obtained by combining Eqs. (5.31), (5.32), and (5.33)

$$D^* = \frac{\sqrt{A_d\,\Delta f}}{\Phi_{det}} SNR \tag{5.35}$$

$$D^* = \frac{\sqrt{A_d\,\Delta f}}{v_{n,rms}} \mathcal{R}_v \tag{5.36}$$

www.EngineeringEbooksPdf.com

$$D^* = \frac{\sqrt{A_d \, \Delta f}}{i_{n,rms}} \mathcal{R}_i \,, \tag{5.37}$$

where we can see that D* is proportional to SNR and also to responsivity. Recalling from Eq. (5.33) that D* is specified in energy-based units, we find that when D* is plotted as a function of wavelength for an ideal photon detector (Fig. 5.17), a curve shape similar to Fig. 5.3 is seen, with a linear increase in D* with wavelength up to λ_{cutoff}. Spectral D* for an ideal thermal detector is flat with wavelength, similar to Fig. 5.1.

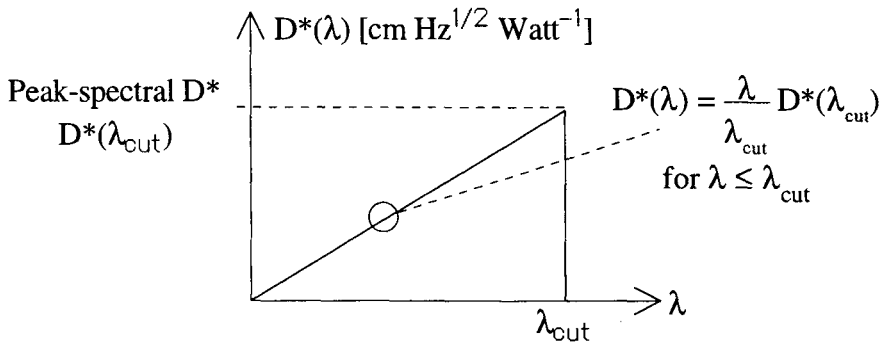


Figure 5.17. D as a function of wavelength for an ideal photon detector.*

We now consider some numerical-calculation examples using D*, remembering that when the detector area and bandwidth change, NEP changes but D* does not. A detector of dimensions 1 mm × 1 mm has a responsivity of 100 V/W, and is measured to have 3 μV of rms noise within a 1-MHz bandwidth. Using Eq. (5.36) we calculate D* as 3.3×10^9 cm $Hz^{1/2}$ W^{-1}. If this same detector material is fabricated into a 50 μm × 50 μm detector, which is used in a bandwidth of 500 kHz, we can find NEP using Eq. (5.33), given that the D* is independent of area and bandwidth. The NEP is 1.06×10^{-9} W. If 0.5 μW of power is received by the detector, the SNR = 471, according to Eq. (5.31).

5.11 Photovoltaic Detectors

A photovoltaic (PV) detector operates by the mechanism of an absorbed photon generating a hole-electron pair at a p-n junction in a semiconductor. As seen in Fig. 5.18, a built-in electric field exists in the vicinity of the junction. This causes an immediate separation of the hole and the electron once they are generated, allowing the photovoltaic detector to develop a voltage across an open circuit. By comparison, a photoconductor has no junction and thus no built-in electric field. A photoconductor must be biased to sense a change in conductivity, and cannot develop an open-circuit voltage. Being a photon detector, a PV has a cutoff wavelength consistent with Eq. (5.2). Some typical PV cutoff wavelengths are shown in Table 5.2.

www.EngineeringEbooksPdf.com

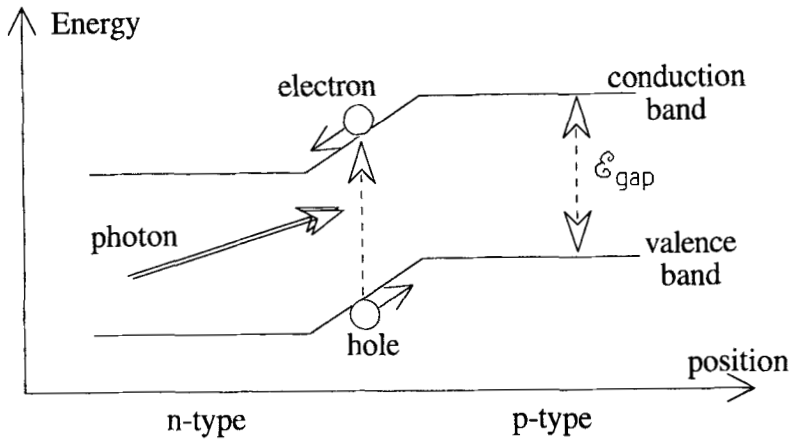


Figure 5.18. Photovoltaic detector.

Table 5.2. Typical cutoff wavelengths for PV detectors.

PV Material	λ_{cut} [μm]
CdS	0.52
CdSe	0.69
GaAs	0.80
Si	1.1
GaAlAs	1.3
Ge	1.9
PbS	3.0
InSb	5.6
$Hg_{1-x}Cd_xTe$	10 – 12 (depending on mixing ratio, x)

The p-n junction in the photovoltaic detector is a diode, so the current-voltage (i-v) characteristic of the PV detector is that of a diode. For the case of no photon irradiance, where the detector is in the dark:

$$I_{dark} = I_0 \left(e^{qV/\beta kT} - 1 \right), \tag{5.38}$$

where β is the diode non-ideality factor, q is the electronic charge, and k is Boltzmann's constant. With photons incident on the detector, the total diode current becomes

$$I_{total} = I_{dark} - I_{photogenerated}, \tag{5.39}$$

where the negative sign comes from the direction of the photogenerated current compared to the convention for positive current flow in a diode. The expression for photogenerated current is

www.EngineeringEbooksPdf.com

$$I_{photogenerated}\left[\frac{coul}{s}\right] = \eta\left[\frac{electron}{photon}\right]\phi_q\left[\frac{photon}{s}\right]q\left[\frac{coul}{electron}\right], \tag{5.40}$$

where η is the quantum efficiency and ϕ_q is the photon flux incident on the detector. A typical i-v plot for a photovoltaic is seen in Fig. 5.19.

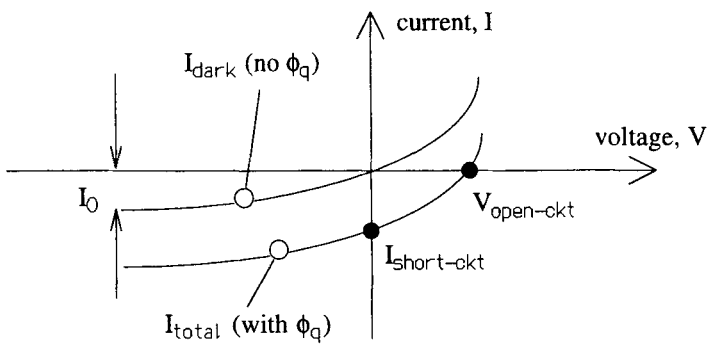


Fig. 5.19. Typical i-v plot for a photovoltaic detector, showing short-circuit and open-circuit operating points.

Circuit interfacing to photovoltaic detectors is typically accomplished using operational amplifiers. In the open-circuit mode of the PV, the total current through the device is held at zero, and photon irradiation produces an open-circuit voltage, as seen in Fig. 5.20. This mode of operation is unique to the PV and cannot be used in a photoconductive detector, which lacks the p-n junction necessary for charge separation.

The magnitude of the open-circuit voltage is found from Eqs. (5.38) - (5.40) as

$$V_{open-ckt} = \frac{\beta kT}{q} Ln\left[\frac{\eta\phi_q q}{I_0} + 1\right]. \tag{5.41}$$

The open-circuit voltage is seen to be proportional to the natural log of the photon flux, providing an increased dynamic range at the expense of linearity.

In the short-circuit mode of operation, the current flow through the diode is measured when the two terminals of the diode are kept at the same voltage, as seen in Fig. 5.21. A short-circuit current is seen to be linearly proportional to the photon flux

$$I_{short-ckt} = -\eta\phi_q q . \tag{5.42}$$

As an example of calculations for short-circuit PV operation, suppose we have a photon flux of 8×10^{12} photons/s. Assuming a quantum efficiency of 80%, we have a photogenerated current of 1.025 μA. If the feedback resistance is 10 kΩ, a voltage of 10.25 mV is seen at the output of the op-amp circuit.

www.EngineeringEbooksPdf.com

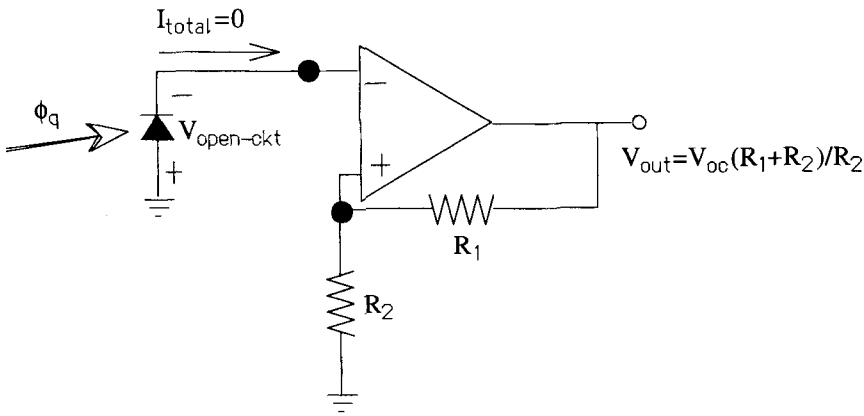


Figure 5.20. Open-circuit operation of a photovoltaic detector.

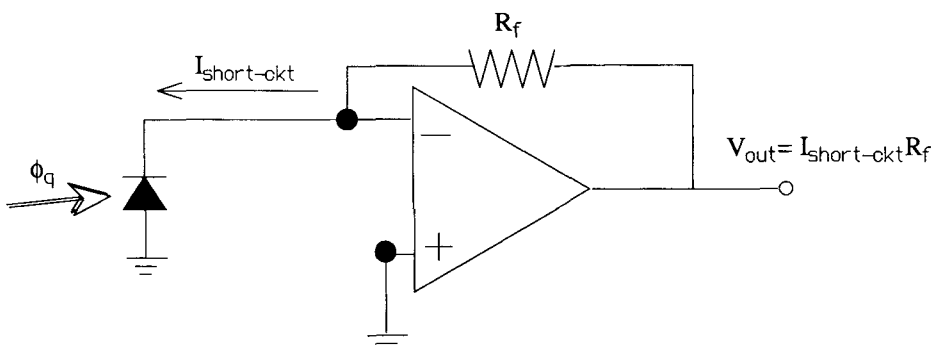


Figure 5.21. Short-circuit operation of a photovoltaic detector.

Photovoltaic detectors can also be operated in a reverse-bias mode, as seen in Fig. 5.22. In this case the current is also linearly proportional to photon flux,

$$I_{short-ckt} = I_0 - \eta\phi_q q \ . \tag{5.43}$$

The advantage of this configuration is that the reverse bias decreases the capacitance of the p-n junction by increasing the separation of the space-charge layer. This is similar to the approach used in p-i-n detectors, where an undoped intrinsic layer accomplishes the same purpose. A decreased capacitance lowers the RC time constant, leading to a faster response.

www.EngineeringEbooksPdf.com

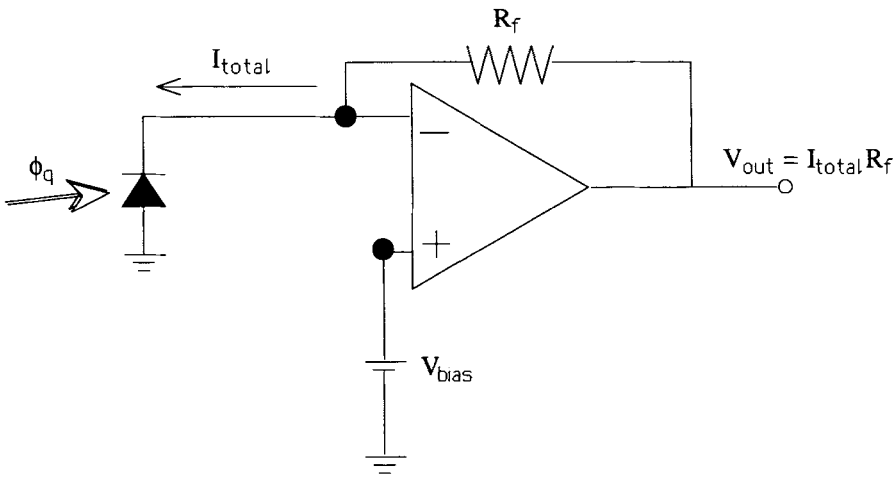


Figure 5.22. Reverse-bias operation of a photovoltaic detector.

Photovoltaic response in the visible and near-IR portion of the spectrum is typically provided by Si detectors, in the 3- to 5-μm band by InSb, and in the 8- to 12-μm band by HgCdTe. Noise-equivalent powers are in the range of 10^{-10} to 10^{-12} W for PV detectors, and time constants are typically in the range of μs to ns, with p-i-n diodes providing the fastest response of the PVs at about 100 ps.

5.12 Schottky-Barrier Detectors

Schottky-barrier detectors, seen in Fig. 5.23, are fabricated of thin (≈ 100 Å) films of metal deposited onto a semiconductor substrate. Silicon is the most commonly used substrate material because of the well-developed fabrication processes available for Si. The metal-semiconductor junction is back-illuminated through the Si, with absorbed photons releasing electrons from the metal into the substrate.

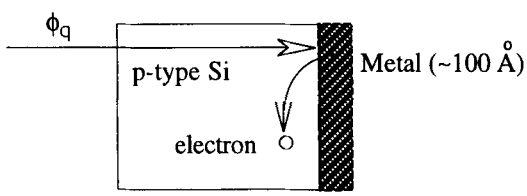


Figure 5.23. Schottky-barrier photdiode.

The quantum efficiency of the photogeneration process is an order of magnitude lower than for photovoltaic materials such as Si and HgCdTe, but the uniformity of response is independent of the exact thickness of the metal. Schottky-barrier detectors can be fabricated in arrays that have excellent detector-to-detector uniformity, which is a valuable feature in IR focal plane arrays. The limiting noise in IR-detector arrays often results from spatial nonuniformity rather than temporal noise.

www.EngineeringEbooksPdf.com

The most common metals used for Si-based Schottky-barrier detectors are: iridium, with a cutoff wavelength of 9.5 μm; platinum, with a cutoff wavelength of 5.6 μm; palladium, with a cutoff wavelength of 2.6 μm; and nickel, with a cutoff wavelength of 1.8 μm. Of these, PtSi is the best-developed technology, providing detector arrays in the 3- to 5-μm band that operate at 77 K.

5.13 Photoconductive Detectors

The mechanism for a photoconductive (PC) detector is the generation of electron-hole pairs by photon absorption. However, unlike a photovoltaic detector, a photoconductor is constructed from homogeneous semiconductor material. Lacking a junction, a photoconductor simply changes conductivity in response to applied photons, and cannot generate an open-circuit voltage. A PC detector must have a bias voltage applied to render the change in conductivity measurable, as seen in Fig. 5.24.

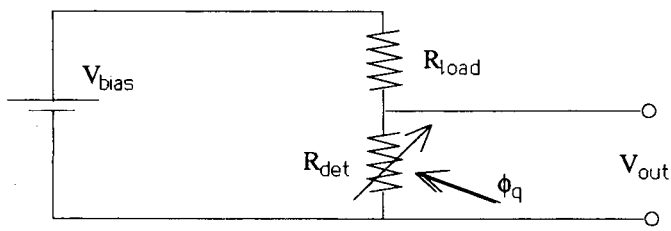


Figure 5.24. Photoconductor readout circuit.

The spectral response of a PC is similar to that of a PV detector, in that the photon to be detected must have sufficient energy to raise an electron across the energy gap from the valence band to the conduction band. In the 3- to 5-μm band, lead salts are important photoconductors, with PbS having a cutoff wavelength of 3 μm, and PbSe having a cutoff of 4.5 μm. Alloys of HgCdTe can be used for PCs that allow cutoff wavelengths out as far as 25 μm. However, because there is no need to create a p-n junction, PC materials have an added design flexibility compared to PV materials. It is possible to take an intrinsic semiconductor and add impurity doping to create an intermediate energy level inside of the energy gap. This configuration is called an extrinsic semiconductor. Typically the host semiconductor for these materials is Ge or Si because of the processing convenience. As seen in Fig. 5.25, the energy gap of the extrinsic material is smaller than the corresponding intrinsic material, and hence the extrinsic PC has a longer cutoff wavelength. Spectral response to 40 μm can be achieved. The two most common extrinsics are Si:In, with a cutoff of 7 μm and Ge:Hg with a cutoff of 14 μm. Longer cutoff wavelengths imply colder operation to suppress dark current, which typically requires liquid He cooling for wavelengths past 12 μm. Extrinsic PCs have an order of magnitude lower quantum efficiency than do intrinsic PCs, becuase the dopant levels in a semiconductor lattice are necessarily much lower than the carrier concentrations arising from the host atoms in an intrinsic material.

Photoconductor performance in terms of NEP is similar to that of the corresponding PV detectors, except that photoconductors have a noise penalty of the square root of two because generation-recombination noise has a higher rms variation than does shot noise. In addition, 1/f noise is always present in a PC, because there must be a dc current flow. NEPs of 10^{-10} to 10^{-12} W are typical.

www.EngineeringEbooksPdf.com

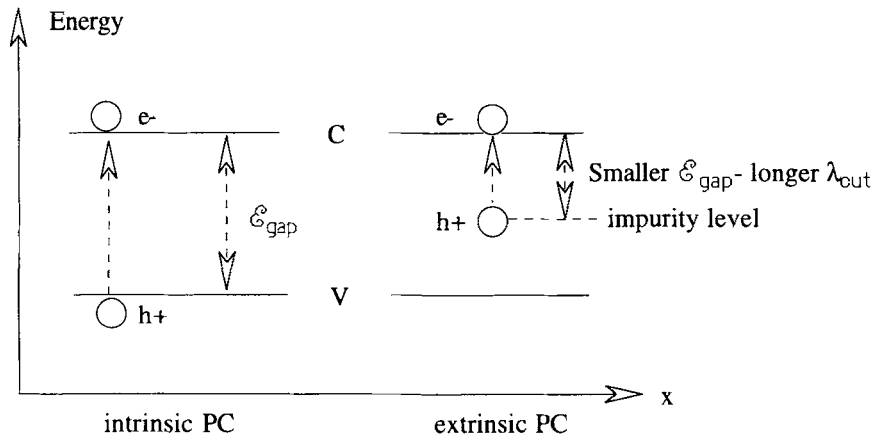


Figure 5.25. Comparison of intrinsic and extrinsic energy-band diagrams.

Time constants for PCs are also similar to those for PVs, typically in the μs to ns range. Photoconductive gain (a phenomenon where the photogenerated carrier lifetime is longer than the carrier transit time through the device) can be used to increase the signal-to-noise ratio, at the expense of a slower temporal response.

5.14 Photoemissive Detectors

Another class of photon detector operates on the photoelectric effect. Electrons are emitted into a vacuum from a metal surface called a cathode when photons are absorbed. The photons must have sufficient energy to overcome the "work function" of the metal surface. Typical photoemissive detectors have their best response in the high-energy (blue) end of the spectrum, with limited response toward the red and near IR. The photoexcited electrons are accelerated across a vacuum gap by application of high (typically kV) voltage and are collected on a metal plate called an anode. The basic form of a photoemissive detector is seen in Fig. 5.26. Because of the requirements of an optically transparent evacuated enclosure and a high voltage, this type of detector is not practically used in the form seen in Fig. 5.26, but is used only in conjunction with electron-amplification techniques that operate by secondary emission of electrons.

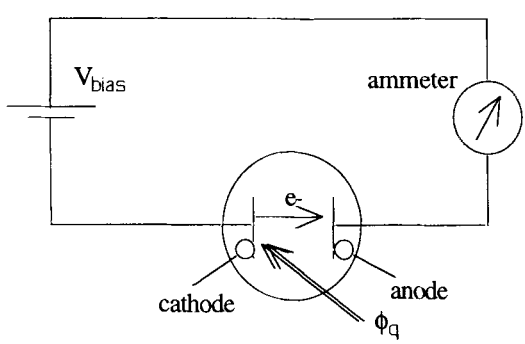


Figure 5.26. Vacuum photodiode.

www.EngineeringEbooksPdf.com

Secondary emission of electrons allows construction of a high-gain, low-noise current amplifier. Figure 5.27 is a schematic of a photomultiplier tube (PMT), which represents the best sensitivity obtainable from a photon detector. Electrons that are accelerated in a vacuum are allowed to impinge on a surface called a dynode. When a single electron reaches the dynode surface, the impact of the electron releases more electrons from the surface. If several dynodes are used in series, with an electron-accelerating voltage across each, a substantial current can be collected at the anode in response to the absorption of only a single electron at the cathode. The amplification of the photocurrent in this way has noise advantage over taking the photocurrent produced in the vacuum photodiode of Fig. 5.26 directly to an external amplifier.

Performance of PMTs is generally limited by shot noise of either the signal current or the dark current. Dark current is caused by thermionic emission from the cathode and can be decreased by cooling the sensor. Response to individual photon events is possible, with a NEP in the range of 5×10^{-19} W. Responsivity is best in the blue, with some capability to 1.2 μm with "negative electron affinity" cathode materials, which have a particularly low work function. Time response for PMTs can be as short as 1 ns. For certain laboratory-based applications, the advantages of single-photon sensitivity (allowing photon counting) and fast response time can outweigh the disadvantages of the PMTs limited spectral response, and the requirements of an evacuated glass envelope and a high-voltage bias source.

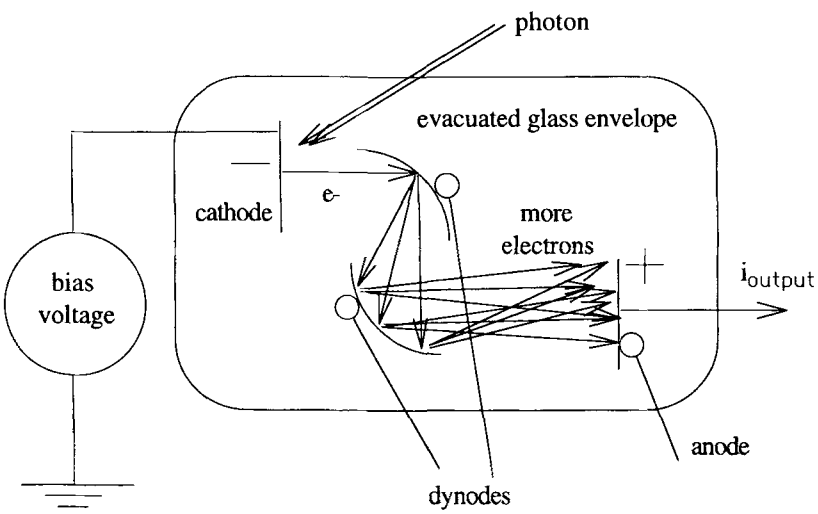


Figure 5.27. Photomultiplier tube.

5.15 Bolometric Detectors

The bolometric sensor is a common type of thermal detector. A bolometer is a resistor with a temperature-dependent resistance. Photons are absorbed on the sensor surface, and the energy of the absorbed photons cause a temperature rise in the sensor. The resulting change in resistance is sensed by an external circuit, as seen in Fig. 5.28. As with the photoconductor, a bias current is required to sense the change in resistance. The load resistor is often an element identical to the sensor but shielded from radiation. This allows the circuit to be insensitive to changes in ambient temperature. As with the

www.EngineeringEbooksPdf.com

photoconductor, a bias current is required to sense the change in resistance. The load resistor is often an element identical to the sensor but shielded from radiation. This allows the circuit to be insensitive to changes in ambient temperature.

A figure of merit for bolometers is α, the fractional change in resistance per degree of temperature change

$$\alpha \equiv \frac{1}{R}\frac{\partial R}{\partial T} . \tag{5.44}$$

Typically α is positive ($\approx 0.5\%/°C$) for metals, because resistance of a metal rises with increased temperature. For superconducting materials held near the transition temperature, α is also positive and is orders of magnitude larger. For semiconductors, the resistance decreases with increasing temperature, giving α a value in the range of $-5\%/°C$ for semiconductors at room temperature. Typical bolometer resistances are 10 kΩ to 1 MΩ.

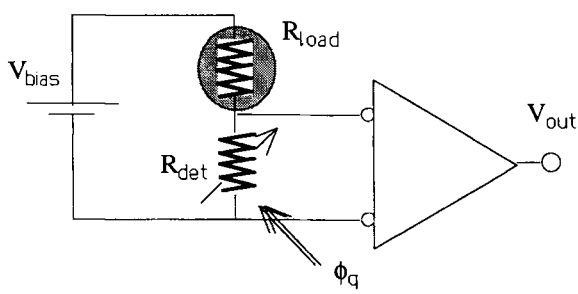


Figure 5.28. Bolometric readout circuit.

The primary advantage of bolometers is their wide spectral response, from visible to the long-wave IR. They are useful in applications where cryogenic cooling is not feasible (from a weight, power, or cost viewpoint) but where IR detection is required at wavelengths that would require photon detectors to be cooled. Time constants are typically in the range of 1 to 100 ms, with NEPs in the range of 10^{-8} to 10^{-10} W. Better sensitivities are obtained for the longer time constants. Bolometers have been recently demonstrated in large focal-plane-array configurations, allowing development of uncooled IR imaging sensors.

5.16 Pyroelectric Detectors

Another useful thermal detector is the pyroelectric. These sensors are fabricated from materials (such as triglycerine sulfate) of sufficient asymmetry that they possess a permanent dipole moment even in the absence of an applied electric field. The magnitude of this dipole moment is temperature dependent. The sensing mechanism is based on the fact that as photons are absorbed, the temperature of the element is changed, and there is motion of bound charge on the surface of the material corresponding to the change in dipole moment. If this material is placed between the plates of a capacitor, as seen in Fig. 5.29, the motion of bound charge induces a current flow in the circuit connected to the plates in response to dT/dt. The magnitude of the current flow is

www.EngineeringEbooksPdf.com

$$i = \rho A \frac{dT}{dt}, \tag{5.45}$$

where ρ is the pyroelectric coefficient (in the range of 10^{-8} Coul cm^{-2} K^{-1}) and A is the area of the capacitor. It is seen from Eq. (5.45) that a pyroelectric detector responds only to a change in temperature. Cooling a pyroelectric sensor is generally not required, because there is no dark-current contribution from the ambient-temperature environment.

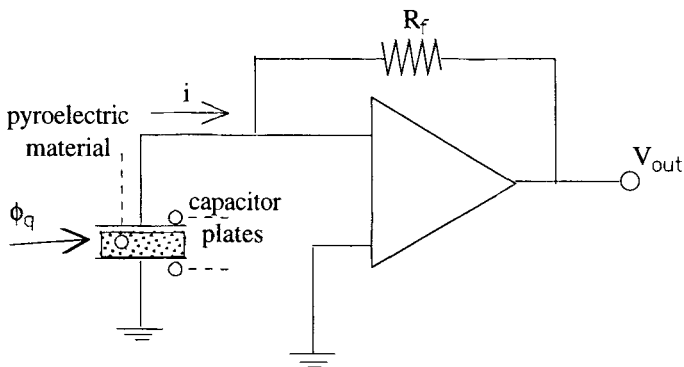


Figure 5.29. Pyroelectric sensor configuration.

Pyroelectric sensors have both an electrical and a thermal time constant. The electrical time constant is just the product of the feedback resistance and the capacitance of the plates. An equivalent circuit used to determine the electrical time constant is shown in Fig. 5.30. The thermal time constant is determined by the thermal mass of the sensor material, and by the thermal conductance of the heat-loss paths.

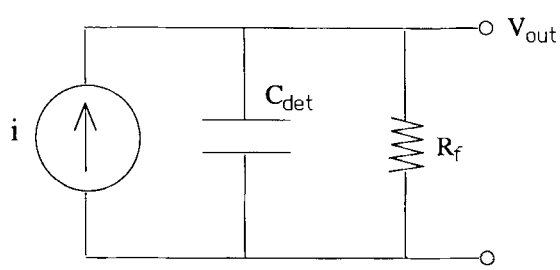


Figure 5.30. Equivalent circuit for pyroelectric sensor.

The value chosen for the feedback resistance R_f affects both the responsivity and the response time. As seen in Fig. 5.31, a large R_f (e.g., 10^9 Ω) gives slow response but large responsivity. Conversely, a small R_f (e.g., 10^3 Ω) gives faster response but at the expense of a smaller responsivity.

www.EngineeringEbooksPdf.com

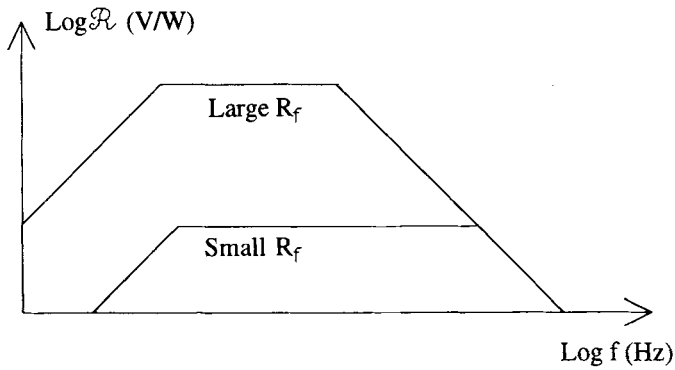


Figure 5.31. Choice of feedback resistor affects both responsivity and response time.

Noise performance for pyroelectrics is determined by Johnson noise and amplifier noise. Better noise performance is achieved by pyroelectrics than by most other uncooled thermal detectors, in the range of 10^{-9} to 10^{-10} W. The time response is quite fast for a thermal detector, with time constants of 100 ns or shorter possible, although with relatively small responsivity. For many laser-measurement applications, the disadvantage of small responsivity is outweighed by the good values of time response and NEP, along with the broad spectral response typical of a thermal detector. Pyroelectric sensors are widely used in laboratory instrumentation.

Bibliography

Dereniak, E., and G. Boreman, *Infrared Detectors and Systems*, Wiley, New York, 1996.
Keyes, R., *Optical and Infrared Detectors*, Springer-Verlag, New York, 1980.
Boyd, R., *Radiometry and the Detection of Optical Radiation*, Wiley, New York, 1983.
Wolfe, W., and G. Zissis, *The Infrared Handbook*, Environmental Research Institute of Michigan, Ann Arbor, 1978.

www.EngineeringEbooksPdf.com

Copyrighted Materials
Copyright © 1998 SPIE Retrieved from www.knovel.com

Chapter 6
Lasers

6.1 Introduction

Our discussion of lasers will concentrate on the unique properties of laser radiation. We will initially consider the characteristics of directionality, monochromaticity, and brightness, and then investigate the Gaussian wave nature of a typical laser output beam. We will conclude with issues relating to measurement of laser-beam parameters.

6.2 Directionality, Monochromaticity, and Brightness

Directionality is one of the most useful properties of laser radiation. Laser light propagates as a highly collimated beam, and has a diffraction angle that is determined soley by diffraction effects. This high degree of collimation is unique to lasers, and is a consequence of the fact that a laser is an extended source of plane wavefronts, as seen in Fig. 6.1. The entire exit aperture of a laser emits rays that are parallel to the optic axis. This behavior can be compared to a nonlaser extended source, where each point on the source is an independent radiator that emits rays in all forward directions.

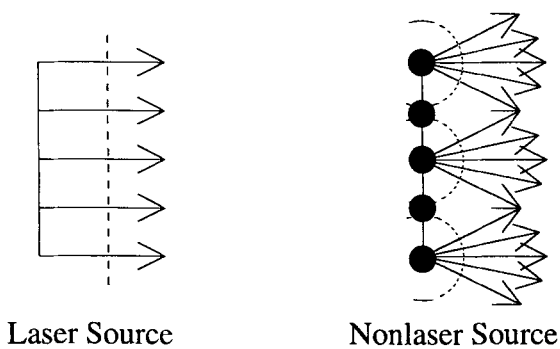


Figure 6.1. Comparison of laser and nonlaser extended sources.

Each source will have a divergence angle caused by diffraction, which is approximately the wavelength divided by the beam diameter

$$\theta \approx \lambda / D_{beam} \quad . \tag{6.1}$$

In a nonlaser source, the beam divergence caused by the lack of collimation is much larger than that caused by diffraction. In the laser source, the diffraction spreading is the only contribution to beam divergence.

A laser beam can be focused to a very small spot, whose size is determined by diffraction. Only a very small region of the surface of the nonlaser source can be focused to a diffraction-limited spot, as seen in Fig. 6.2. The size of that region is, as seen in Fig. 1.26, determined by the footprint of the diffraction blur in the object plane of the

www.EngineeringEbooksPdf.com

optical system used to focus the radiation. Using such a small region of the nonlaser source does not yield much focusable power. For example, the amount of flux that passes through a 20-μm pinhole placed in front of a blackbody source is negligible.

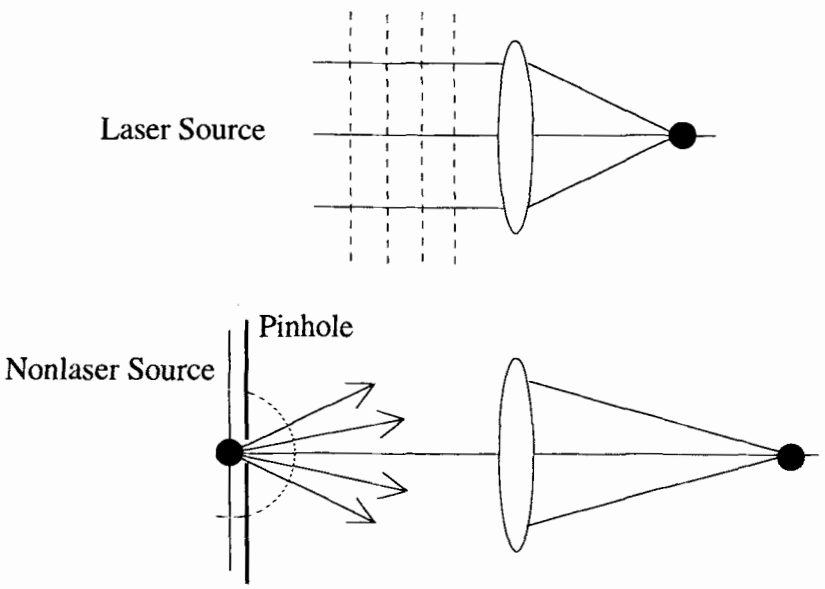


Figure 6.2. Focusing laser and nonlaser sources.

To obtain more focused power from a nonlaser source, we must increase the area of the source that contributes to the focused beam (Fig. 6.3). In this case, an image of an extended portion of the source is produced, with a size determined by geometrical magnification effects, rather than a diffraction-limited spot. The issue of interest for applications such as drilling or welding is spatial power density rather than power. The total power in the focused image is increased, but the irradiance remains constant.

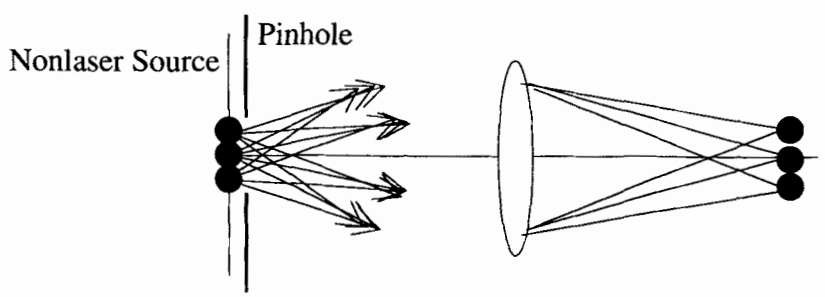


Figure 6.3. Focusing of an extended region of a nonlaser source.

www.EngineeringEbooksPdf.com

Monochromaticity is another unique property of laser radiation. Lasers have a very narrow spectral bandwidth $\Delta\lambda$

$$\lambda_{laser} = \lambda_{midband} + \Delta\lambda \quad . \tag{6.2}$$

A typical $\Delta\lambda$ is 10^{-5} to 10^{-8} μm. This very small wavelength interval can also be more conveniently and commonly expressed in terms of a frequency interval. From the relationship between frequency and wavelength ($f = c/\lambda$), we obtain by a derivative

$$|\Delta f| = \Delta\lambda \, (c/\lambda^2) \quad . \tag{6.3}$$

For example, a laser with a $\lambda = 0.63$ μm and a $\Delta\lambda = 10^{-5}$ has a frequency bandwidth $\Delta f \approx$ 8000 MHz. A laser at the same center wavelength and $\Delta\lambda = 10^{-8}$ has $\Delta f \approx 8$ MHz.

Compared to nonlaser sources, a laser is perceived to be very bright because the solid angle into which the beam radiates and the spectral bandwidth of the beam are both small. As a means to compare the perceived brightness, let us compare the spectral radiance, L_λ (W cm^{-2} sr^{-1} μm^{-1}), of a 5-mW laser having a divergence angle $\theta = 1$ mrad, and $\Delta\lambda = 10^{-6}$ μm, with the spectral radiance of the sun (approximated by a blackbody at 6000 K) we find that $L_{\lambda,laser} \approx 10^{12}$, while $L_{\lambda,sun} \approx 10^{6}$.

6.3 Gaussian Beams

The simplest output from a laser has a circular Gaussian profile as seen in Fig. 6.4. Expressed in terms of irradiance, a radial profile of the beam has the form

$$E(\rho) = E_0 \exp\{-2\rho^2/w^2\} \, , \tag{6.4}$$

where w is the semidiameter of the beam to the $1/e^2$ point (13.5% of the on-axis value) in irradiance

$$E(\rho=w) = E_0 \exp\{-2w^2/w^2\} = E_0 \exp\{-2\} = E_0 \times 0.135 \quad . \tag{6.5}$$

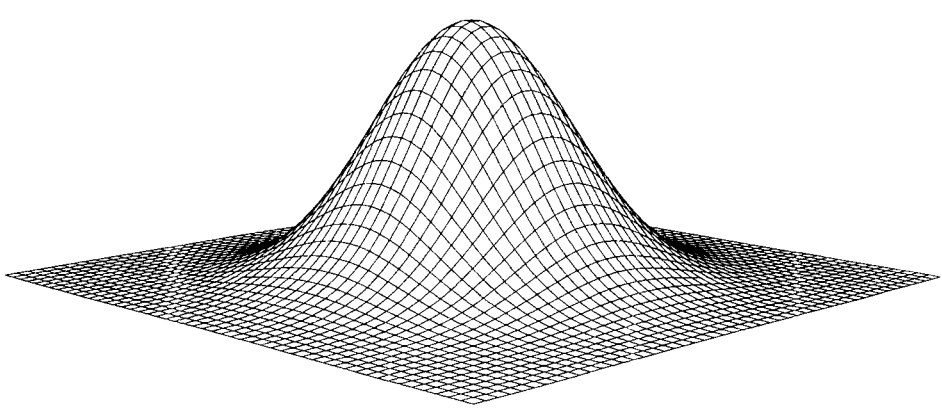


Figure 6.4. Circular Gaussian irradiance profile of a laser beam.

www.EngineeringEbooksPdf.com

As seen in Fig. 6.5, given a value of w, any other beam dimension can be calculated. The full-width-at-half-maximum (FWHM) can be found by solving for ρ such that

$$E(\rho) = 0.5\, E_0 \tag{6.6}$$

$$0.5 = \exp\{-2\rho^2/w^2\} = \mathrm{Ln}\,\{0.5\} = -2\rho^2/w^2 \tag{6.7}$$

$$\rho = 0.589\ w\ ;\ \mathrm{FWHM} = 2\rho = 1.18\ w\ . \tag{6.8}$$

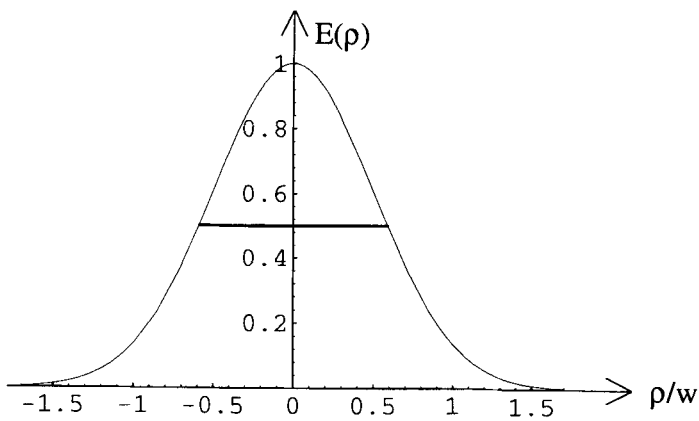


Figure 6.5. FWHM of laser beam is 1.18 w.

Because the irradiance profile of the beam is Gaussian, it extends to infinity in the transverse direction. From the point of view of the size, weight, and cost of the optical system, we must truncate the beam at some finite diameter and will lose some of the power in the beam. The diameter of the optics is typically sized on the following criteria. An aperture of diameter equal to 2w will pass 86% of the beam. An aperture of diameter equal to 3w will pass 99% of the beam. Often, even for those situations in which the power loss can be tolerated, the optics is sized at the D = 3w criterion because, for that aperture size, the amount of diffracted power arising at the edge of the beam being truncated is minimal.

Laser beams having a Gaussian profile will change w as they propagate, but will keep a Gaussian profile for all distances z, as seen in Fig. 6.6. This property is a consequence of diffraction, because the Fourier transform of a Gaussian is also a Gaussian. Figure 6.6 assumes that the attenuation of the propagation medium is zero. The decrease in peak beam irradiance seen with increasing z is a consequence of the fact that the finite beam power is spread over a beam of increasing semidiameter as z increases.

The beam waist, denoted w_0, is the minimum size that the beam attains as a function of z. Diffraction determines the divergence angle θ of a Gaussian beam, which behaves as if the beam diffracts from an effective aperture that is approximately the size of the beam waist

$$\theta \approx \lambda/\pi w_0\ . \tag{6.9}$$

www.EngineeringEbooksPdf.com

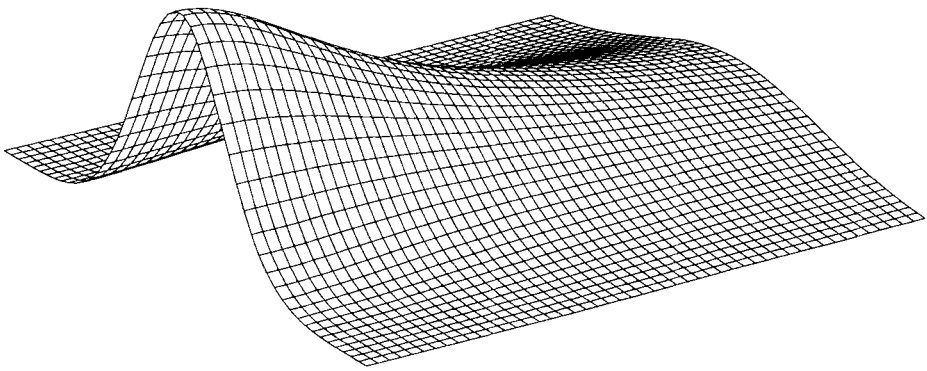


Figure 6.6. Gaussian beam changes semidiameter with z but keeps Gaussian profile.

We see in Fig. 6.7 that a beam with a smaller waist diverges more rapidly and a beam with a larger waist diverges more slowly. The product of beam size and divergence angle is a constant. Beam-expanding telescopes as seen in Fig. 6.8 are often used to increase the beam diameter and hence reduce the divergence angle, especially where beam propagation over a long distance is required. Expansion of the beam before propagation results in a smaller beam size and higher irradiance downstream.

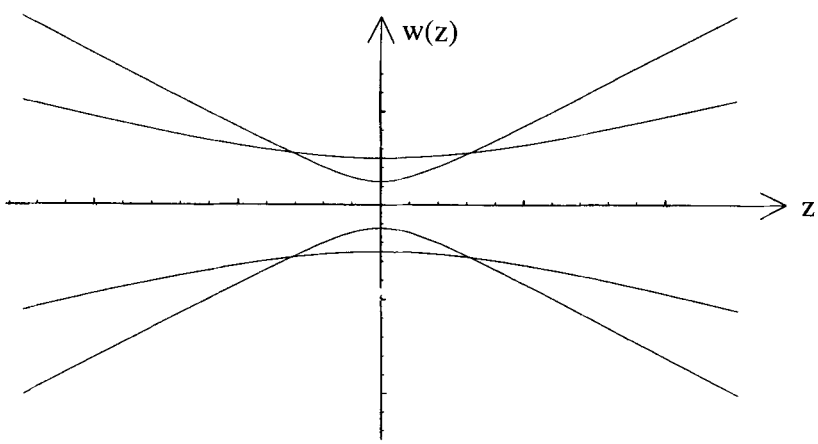


Figure 6.7. Divergence angle related to beam waist size.

www.EngineeringEbooksPdf.com

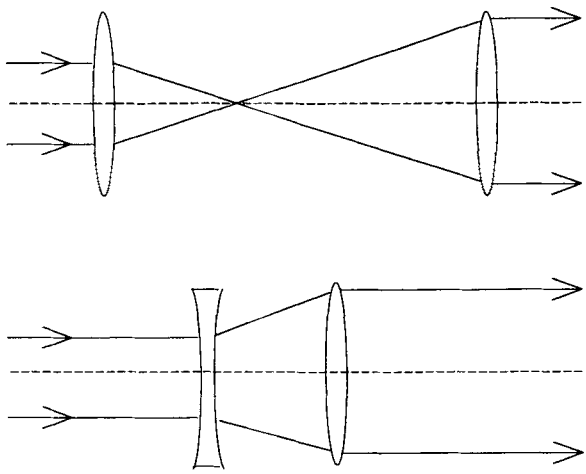


Figure 6.8. Beam-expanding telescopes reduce divergence.

Calculations with Gaussian beams are based on the propagation equation, which defines the beam semidiameter w as a function of distance z. The location of the beam waist is taken to be at the z = 0 point.

$$w(z) = w_0 \sqrt{1 + \left(\frac{\lambda z}{\pi w_0^2}\right)^2} \tag{6.10}$$

We can use Eq. (6.10) to investigate the effect of expansion on long-range beam size. Assuming a wavelength of 0.6 μm, let us compare the beam size w at z = 1 km for a beam with w_0 = 1 mm and for the same beam after passing through a beam expander of the type seen in Fig. 6.8, starting with a w_0 = 5 mm. We find that w(1 km) = 20 cm for the unexpanded beam, and w(1 km) = 4 cm for the expanded beam.

We can find the initial beam-waist size w_0 that minimizes w for any particular distance z (neglecting the effect of atmospheric turbulence) by setting the partial derivative $\partial w(z)/\partial w_0 = 0$.

Evaluating Eq. (6.10) in the limit of large z, such that

$$w\left(z >> \frac{\pi w_0^2}{\lambda}\right) \rightarrow \frac{\lambda z}{\pi w_0} \tag{6.11}$$

we obtain an expression for the far-field divergence angle

$$\theta = \frac{w(z >> \pi w_0^2/\lambda)}{z} = \frac{\lambda}{\pi w_0} \quad . \tag{6.12}$$

www.EngineeringEbooksPdf.com

The Rayleigh range z_R is the distance from the waist at which the semidiameter of the beam has increased to a factor of 1.4 w_0

$$z_R = \pi w_0^2/\lambda \ , \tag{6.13}$$

which can be verified by substitution of Eq. (6.13) into Eq. (6.10). Rayleigh range is a measure of the distance over which the beam remains collimated. Note the quadratic dependence of Rayleigh range on the beam waist size. A larger beam remains collimated over a much longer distance.

Focusing a laser beam to a small spot is important to many applications of lasers. For the typical case seen in Fig. 6.9 of the input beam waist located at the lens, the focused spot size is calculated by multiplying the divergence angle θ of the input beam by the lens focal length. We obtain an expression for the size of the beam waist located at the focus of the lens

$$w_{focus} = \theta\, f = \frac{\lambda\, f}{\pi\, w_{input}} \ . \tag{6.14}$$

According to Eq. (6.14), given light with a wavelength of 0.63 μm, a focal length of 20 mm (corresponding to a 10-power microscope objective), and a w_{input} = 0.5 mm, we find w_{focus} = 8 μm. For the usual lens-diameter criterion used with Gaussian beams (D_{lens} = 3 w_{input}), Eq. (6.14) is consistent with the usual diffraction formula of Eq. (1.21). The difference between the two formulations is that Eq. (1.21) assumes that the aperture of the system is uniformly illuminated, and Eq. (6.14) accounts for the apodization resulting from the Gaussian beam profile. In both cases, the focused spot size (in the absence of aberrations) is determined by diffraction effects.

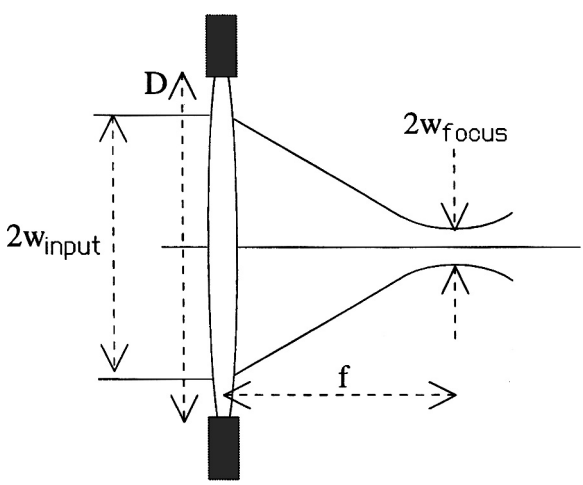


Figure 6.9. Focusing a Gaussian beam to a spot.

www.EngineeringEbooksPdf.com

6.4 Temporal Laser-Beam Measurements

In this section we consider temporal and spatial measurements of laser beams. From the temporal viewpoint, the simplest measurement is average power $\overline{\phi}$ in W. This is adequate for continuous wave (cw) laser output, which does not have a fast time dependence. The output voltage from the detector is proportional to the average power falling on the detector within the measurement interval. A thermal detector is typically used for an average-power measurement, with a time constant in the range of 0.2 to 0.5 s. Typical measurement accuracy is 5%, over a range of beam powers from 100 μW to 100 W. Generally the active detector area is of sufficient size to acquire the entire beam.

For measurement of pulsed laser output, the parameters of interest seen in Fig. 6.10 are $\phi(t)$ (laser power as a function of time), the pulse duration τ (how much time the laser power exceeds a certain threshold value), the period T (amount of time in which a periodic P(t) repeats itself), the repetition rate 1/T (number of pulses per second), and the duty cycle τ/T (fraction of time for which the laser power exceeds threshold). Typical values of duty cycle are 0.1% to 5%. The peak power is the maximum value of $\phi(t)$, which can be measured with a "peak hold" circuit following the detector. The average power is the energy per pulse divided by the period. The energy per pulse (Joules) is the area under the $\phi(t)$ during the pulse duration.

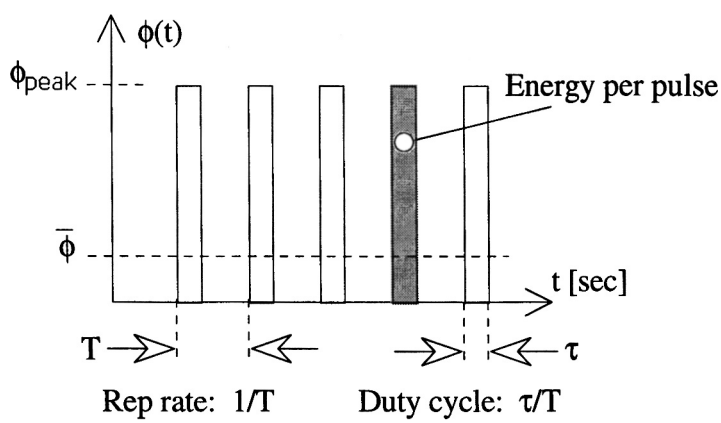


Figure 6.10. Temporal laser-beam parameters.

Given a measurement of average power of 300 mW, with a duty cycle of 1%, and a repetition rate of 40 pulses/s, we can find (assuming square pulses) the energy per pulse and the peak power. The period T = 25 ms, and $\tau = 0.01 \times T = 250$ μs. For 300 mW average power, the energy contained in one period is 300 mW × 0.025 s = 7.5 mJ. However, this energy is contained in 250 μs of the pulse. Thus, the peak power is 7.5 mJ/250 μs = 30 W. The energy per pulse (7.5 mJ) is equal to the energy per period.

6.5 Spatial Laser-Beam Measurements

When measuring the spatial distribution of irradiance of a laser beam, the main parameter of interest is the diameter to the $1/e^2$ point in irradiance (2w). Also of interest is the spatial profile function: that is, how closely is the beam modelled by a Gaussian?

There are three approaches to this problem. The beam size can be estimated by

www.EngineeringEbooksPdf.com

eye with a calibrated target. An example of this is to shine the beam on the wall and measure the width with a ruler. This approach is cheap and convenient, but it has poor accuracy, and is limited in effectiveness to relatively large beams. A second method is to use an array of individual photodetectors, such as a CCD array. This approach is fairly expensive, but has the advantage of being able to display the beam profile on a TV monitor or an oscilloscope in real time. The beam size must be fairly large for this approach to work well. Considering that the detector-to-detector spacing is generally greater that 15 μm, this technique is not appropriate for tightly focused beams.

The most versatile technique is a movable blocking device with a single large detector. This method is inexpensive to implement, and can measure the profile of beams near their focus. As seen in Fig. 6.11, typically a blocking element (pinhole, slit, or knife edge) is placed on a movable micrometer stage in front of the detector. The output is recorded as a function of the along-scan position of the blocking device. Multiple scans can be recorded at various along-beam positions to find the best-focus position of the beam.

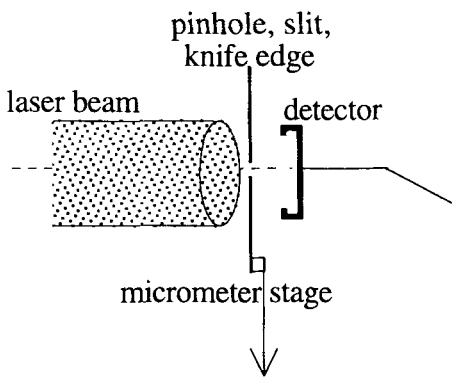


Figure 6.11. Movable blocking device for beam-profile measurement.

When a pinhole is used as the blocking device, a point measurement of the beam is approximated (Fig. 6.12). There is a small amount of averaging in the perpendicular-to-scan direction. A pinhole scan provides information on the actual profile of the beam, and can verify or disprove the assumption of a Gaussian profile. The pinhole measurement does not require a large dynamic range in the detector, in that the whole beam is never allowed to fall on the detector. Because the measured profile will be the convolution of the beam profile with the pinhole profile, the pinhole size must be considered relative to the size of the beam to be measured. The effect of the finite dimension of the pinhole can be removed by a deconvolution procedure in the data processing.

Using a slit, as seen in Fig. 6.13, rather than a pinhole results in averaging of the beam profile in the direction perpendicular to the scan. This yields less information about the actual profile than the pinhole measurement. The resulting data will usually approximate a Gaussian, even if the original beam profile is not. Compared to a pinhole scan, a larger linear dynamic range is required of the detector system, because more of the beam is passed to the sensor. As with pinhole-based measurements, the effect of the slit width must also be considered in the interpretation of the measurement results.

www.EngineeringEbooksPdf.com

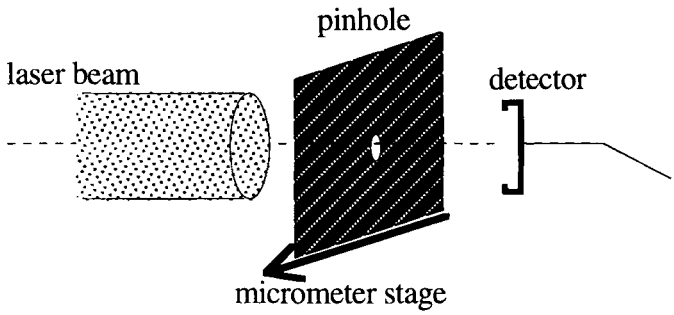


Figure 6.12. Beam-profile measurement using a pinhole.

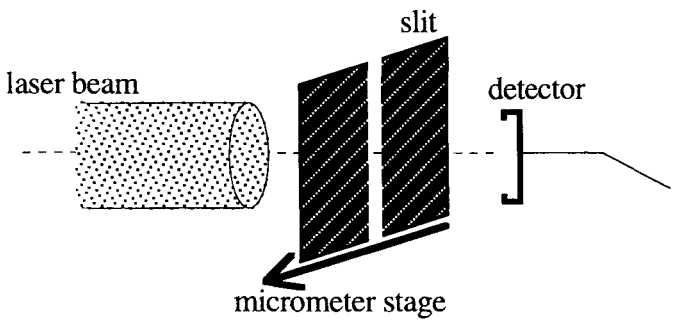


Figure 6.13. Beam-profile measurement using a slit.

If a knife edge (Fig. 6.14) is used as the beam-blocking device there is no convolution effect in the data. The knife-edge scan produces a cumulative profile of the beam, averaged in the perpendicular-to-scan direction. As seen in Fig. 6.15, taking the spatial derivative of the knife-edge data yields the beam profile. A good initial signal-to-noise ratio is required because the derivative process is susceptible to noise in the data. This approach requires the widest linear dynamic range from the detector, because the amount of flux passing to the sensor ranges from zero to the full beam power.

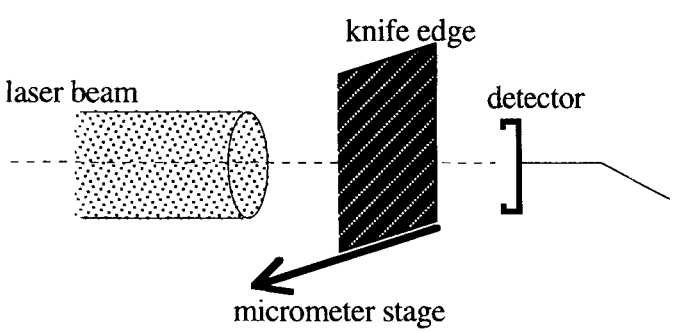


Figure 6.14. Beam-profile measurement using a knife edge.

www.EngineeringEbooksPdf.com

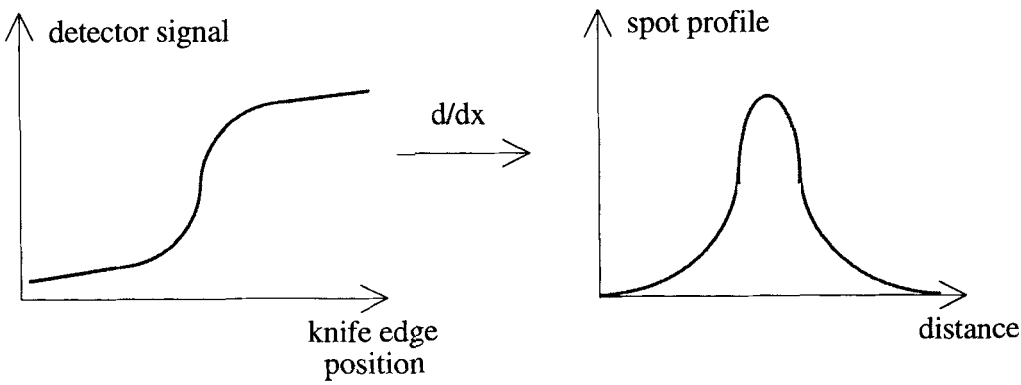


Figure 6.15. Data processing for knife-edge scan.

Figure 6.16 illustrates a convenient way of displaying a laser beam profile in real time, using a stationary slit and detector, with a beam that is moved past them by a galvanometer mirror or a rotating-polygon scanner. This method requires distance for implementation of the optical lever, and so will not work close to a focusing lens. The beam scans repetetively across the detector, so that the results can be displayed on an oscilloscope, which is useful if the beam changes with time.

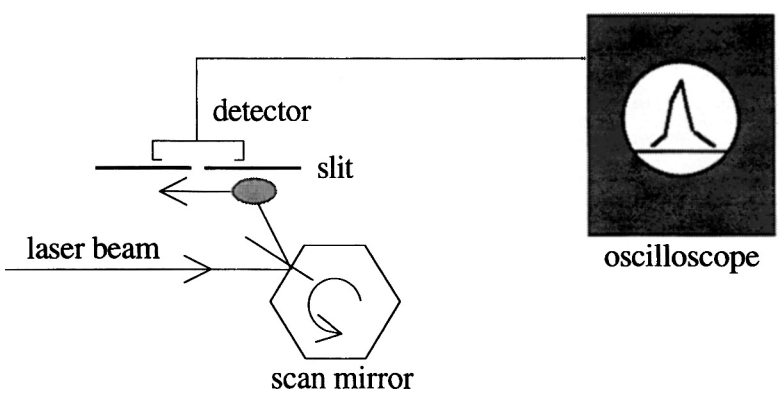


Figure 6.16. Beam-profile measurement using a stationary slit and a scanned beam.

Bibliography

O'Shea, D., W. Callen, and W. Rhodes, *An Introduction to Lasers and Their Applications*, Addison-Wesley, Reading, 1977.
Siegman, A., *Lasers*, University Science Books, Mill Valley, 1986.
Svelto, O., *Principles of Lasers*, Plenum, New York, 1989.
Verdeyen, J., *Laser Electronics*, Prentice-Hall, Englewood Cliffs, 1989.
Yariv, A., *Optical Electronics*, Holt, New York, 1985.

www.EngineeringEbooksPdf.com

Copyrighted Materials
Copyright © 1998 SPIE Retrieved from www.knovel.com

INDEX

This page has been reformatted by Knovel to provide easier navigation.

www.EngineeringEbooksPdf.com

This page has been reformatted by Knovel to provide easier navigation.

www.EngineeringEbooksPdf.com

This page has been reformatted by Knovel to provide easier navigation.

www.EngineeringEbooksPdf.com

This page has been reformatted by Knovel to provide easier navigation.

www.EngineeringEbooksPdf.com

This page has been reformatted by Knovel to provide easier navigation.

www.EngineeringEbooksPdf.com

This page has been reformatted by Knovel to provide easier navigation.

www.EngineeringEbooksPdf.com

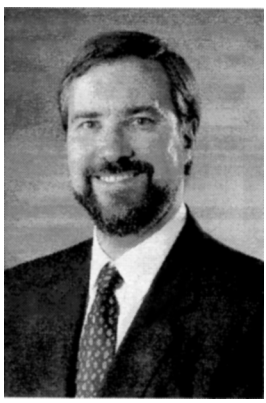
This page has been reformatted by Knovel to provide easier navigation.

www.EngineeringEbooksPdf.com

This page has been reformatted by Knovel to provide easier navigation.

www.EngineeringEbooksPdf.com

Copyrighted Materials
Copyright © 1998 SPIE Retrieved from www.knovel.com



Glenn D. Boreman is Professor of Electrical Engineering in the Center for Research & Education in Optics & Lasers (CREOL) at the University of Central Florida. He received a BS from the Institute of Optics, University of Rochester, and a PhD from the Optical Sciences Center, University of Arizona. He is a registered professional engineer, and has held visiting research positions with IT&T, Texas Instruments, US Army Night Vision Lab, US Army Redstone Arsenal, and McDonnell Douglas. He has been a visiting scholar at the Imperial College in London, and the Swiss Federal Institute of Technology (ETH) in Zürich.

Dr. Boreman serves the Optical Society of America as the Editor of the Optical Technology Division of *Applied Optics*. He also serves on the Board of Directors of SPIE, and is an Associate Editor of the SPIE journal *Optical Engineering*. He is author of the optical-transfer-functions chapter in the *Handbook of Optics*, is coauthor of the graduate textbook *Infrared Detectors and Systems* (John Wiley & Sons, 1996), and has published more than 80 articles in the areas of infrared detectors, statistical optics, and modulation-transfer functions. He received the 1995 Kingslake Medal and Prize from SPIE.

www.EngineeringEbooksPdf.com