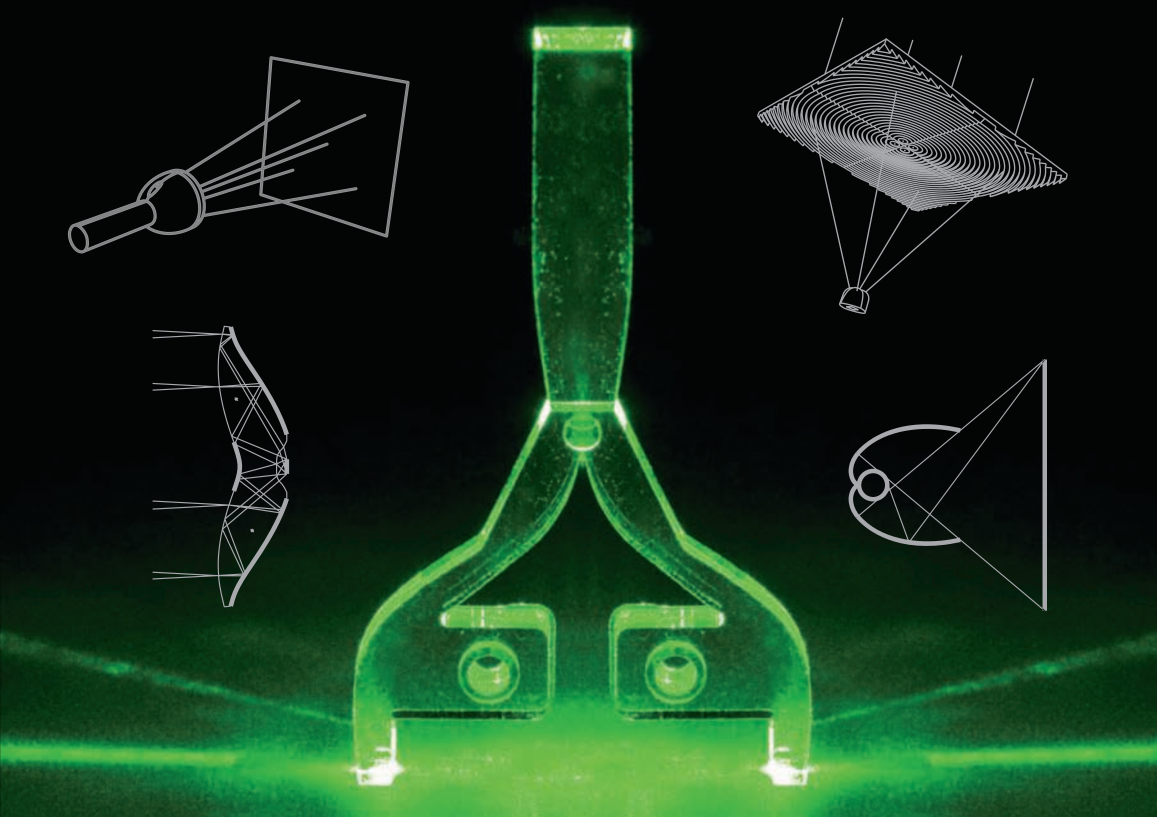
SECOND EDITION

INTRODUCTION TO Nonimaging Optics

JULIO CHAVES



CRC Press
Taylor & Francis Group

SECOND EDITION

INTRODUCTION TO Nonimaging Optics

SECOND EDITION

INTRODUCTION TO Nonimaging Optics

JULIO CHAVES

LIGHT PRESCRIPTIONS INNOVATORS
MADRID, SPAIN

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20150130

International Standard Book Number-13: 978-1-4822-0674-6 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface

Over the past few years some significant nonimaging optical devices have been developed. The second edition of *Introduction to Nonimaging Optics* reflects these developments. Both solar energy concentration and LED illumination have benefited from these late developments. In particular, Köhler illumination combined with nonimaging optics methods led to new optics that address some of the challenges in these fields.

This new edition includes about 45% new material in four new chapters and additions to existing chapters. As with the first edition, it assumes no previous knowledge of nonimaging optics and now covers a wider range of subjects. Some chapters contain intuitive descriptions of the design methods or reasons to use nonimaging devices, while others delve deeper into the theoretical fundamentals or descriptions of more advanced optics.

New Chapter 1 contains an intuitive description of the advantages of nonimaging optics. It assumes no previous knowledge of the field and lays out some of its basic concepts and justifications for its use.

Chapter 9 was extended to include 3-D free-form optics. These more complex designs have more degrees of freedom, which allows them to be used in more challenging situations. For that reason, 3-D free-form optics are a new important trend in optical design.

New Chapter 10 describes some methods for generating output wavefronts used in the design of optics for prescribed output (intensity or irradiance), a very common problem in optical design. These wavefronts are then combined with nonimaging configurations to obtain optical devices. Some illustrative examples are given, but the same methods may be applied to other situations.

New Chapter 11 describes the limit case in which the étendue of the radiation crossing an optic goes to zero (becomes infinitesimal). Although more limited when compared to SMS optics (described in Chapter 9), these infinitesimal étendue optics are much easier to design. These types of optics have found application mainly in the high concentration of solar energy since the angular aperture of direct sunlight is very small.

New Chapter 12 describes Köhler optics combined with nonimaging methods. This is a powerful combination with many uses. A main application with increasing importance is LED color mixing. New lamps coming to market try to combine a tunable emission spectrum with a prescribed output pattern in compact and efficient devices. Köhler optics by themselves or used in combination with other nonimaging optics devices are a serious contender with good potential in this new trend.

When combined, the SMS prescribed intensity wavefronts and Köhler configurations constitute powerful design tools that can address challenging illumination design problems.

Part II of the book includes new material on integral invariants and their applications to nonimaging optics. Chapter 14 describes some of the theoretical aspects, which are common to other fields such as classical mechanics or analytical dynamics. Chapter 18 applies these concepts to optics, in particular to étendue 2-D, one of the invariants used in nonimaging optics. Section 18.4 also derives the expressions for étendue 2-D, but from a geometrical point of view, without having to rely on the Hamiltonian theory. The remainder of Chapter 18 provides examples of applications.

Julio Chaves

Preface to the First Edition

This book is an introduction to nonimaging optics or anidolic optics. The term *nonimaging* comes from the fact that these optics do not form an image of an object, they are nonimaging. The word *anidolic* comes from the Greek (an+eidolon) and has the same meaning. The words *anidolico/anidolica* are mostly used in the Latin languages, such as Spanish, Portuguese, or French, whereas nonimaging is more commonly used in English.

Many optical systems are designed to form the image of an object. In these systems, we have three main components: the object, the optic, and the image formed. The object is considered as being a set of light-emitting points. The optic collects that light (or part of it) and redirects it to an image. The goal of this image is that the rays of light coming out of one point on the object are again concentrated onto a point. Therefore, it is desirable that there be a one-to-one correspondence between the points on the object and those of the image. Only a few "academic" optical systems achieve this perfectly.

Instead, in nonimaging optical systems, in place of an object there is a light source, and the optic is differently designed; and in place of an image there is a receiver. The optic simply transfers the radiation from the source to the receiver, producing a prescribed radiation distribution thereupon.

Although there has been some pioneering work in nonimaging physical optics, nonimaging optics has been developed mostly under the aegis of geometrical optics. Its applications are also based on geometrical optics. Accordingly, this book deals only with nonimaging geometrical optics.

This branch of optics is relatively recent. Its development started in the mid-1960s at three different locations—by V. K. Baranov (Union of Soviet Socialist Republics), Martin Ploke (Germany), and Roland Winston (United States)—and led to the independent origin of the first anidolic concentrators. Among these three earliest works, the one most developed was the American one, resulting in what nonimaging optics is today.

The applications of this field are varied, ranging from astrophysics to particle physics, in solar energy, and in illumination systems. Solar energy was the first substantial big application of nonimaging optics, but recently, illumination has become the major application driving development. These two applications are of prime importance today, as lighting's cost of energy increases, and awareness of its environmental consequences mounts. Nonimaging optics is the ideal tool for designing optimized solar energy collectors and concentrators, which are becoming increasingly important, as we search for alternative and cleaner ways to produce the energy we need. It is also the best tool for designing optimized illumination optics, which engenders more efficient designs and, therefore, lower energy consumption. In addition, with the advent of solid-state lighting, nonimaging optics is clearly the best tool to design the optics to control the light that these devices produce. With the considerable growth that these markets are likely to have in the near future, nonimaging optics will, certainly, become a very important tool.

This book is an introduction to this young branch of optics. It is divided into two sections: The first one deals with nonimaging optics—its main concepts and methods. The second section is a summary of the general concepts of geometrical optics and some other topics. Although the first section is meant to be complete by itself, many general concepts have a different usage in nonimaging optics than in other branches of optics. That is why the second part may be very useful in explaining those concepts from the perspective of nonimaging optics. It is, therefore, a part of the book that the reader can refer to while reading the first section, should some concepts seem obscure or used differently from what he or she is used to.

Julio Chaves

Acknowledgments

This book is the result of many years of studying and designing nonimaging optical devices. Throughout the whole effort my wife Ana stood by my side. This work would have never been possible without her love and dedication through the years, even as the writing took so much of my time away from her.

My parents Julio and Rosa Maria and my brother Alexandre Chaves always tried to provide the best for me and have always been supportive.

I am fortunate to work with extraordinary and talented people who share a passion for nonimaging optics. They have been a source of inspiration and I have discussed with them, through the years, many of the topics covered in the book: My colleagues who have joined Light Prescriptions Innovators, LLC (LPI) over the years, and especially those who, like me, work in optical design: Waqidi Falicoff, Rubén Mohedano, Maikel Hernández, José Blen, and Aleksandra Cvetković. People at the Universidad Politécnica de Madrid (Technical University of Madrid), in particular those with whom I have worked more closely: Pablo Benítez and Juan Carlos Miñano and also people on their team, Pablo Zamora, Dejan Grabovičkić, and Marina Buljan. Manuel Collares-Pereira and Diogo Canavarro, first at the Instituto Superior Técnico (Higher Technical Institute) in Lisbon, and then at the University of Évora in Portugal.

Author

Julio Chaves was born in Monção, Portugal. He completed his undergraduate studies in physics engineering at the Instituto Superior Técnico (Higher Technical Institute), Universidade Técnica de Lisboa (Technical University of Lisbon), Lisbon, Portugal in 1995. He received his PhD in physics from the same Institute. Dr. Chaves did postgraduate work in Spain during 2002 at the Solar Energy Institute, Universidad Politécnica de Madrid (Technical University of Madrid). In 2003, he moved to California, and joined Light Prescriptions Innovators, LLC (or LPI). In 2006, he moved back to Madrid, Spain, and since then has been working with LPI.

Dr. Chaves developed the new concepts of stepped flow-line optics and ideal light confinement by caustics (caustics as flow lines). He is the coinventor of several patents, and the coauthor of many papers in the field of nonimaging optics. He participated in the early development of the simultaneous multiple surface design method in three-dimensional geometry.

List of Symbols

$\dot{x}$	Total derivative of $x(t)$ where t is time: $\dot{x}(t) = dx/dt$
x'	Total derivative of $x(y)$ where y is a geometrical quantity: $x'(y) = dx/dy$
∇	Gradient of a scalar function: $\nabla F(x_1, x_2, x_3) = (\partial F/\partial x_1, \partial F/\partial x_2, \partial F/\partial x_3)$
$\nabla \times$	Rotational operator (curl)
$[\mathbf{A}, \mathbf{B}]$	Distance between points **A** and **B**
$[[\mathbf{A}, \mathbf{B}]]$	Optical path length between points **A** and **B**
$\lvert a \rvert$	Absolute value of a
$\lVert \mathbf{v} \rVert$	Magnitude of vector **v**
$\langle Z \rangle$	Average value of Z
(x_1, x_2)	Two-dimensional vector or point with coordinates x_1 and x_2
A	Area A in a three-dimensional system
a	Length a in a two-dimensional system
$\mathbf{c}(\sigma)$	Curve with parameter σ
$F(x, y)$	Function F of parameters x and y
$F_{A1\text{–}A2}$	Shape factor from area A_1 to area A_2
H	Hamiltonian (when light paths are parameterized by coordinate x_3)
i_1, i_2, i_3	Generalized coordinates
$\mathbf{J}$	Vector flux
L	Radiance Lagrangian (in the context of Lagrangian optics)
L_V	Luminance
L^*	Basic radiance: $L^* = L/n^2$
L_V^*	Basic luminance: $L_V^* = L_V/n^2$
$\mathbf{n}$	Unit vector normal to a surface or curve. It has components $\mathbf{n} = (m_1, m_2, m_3)$.
n	Refractive index
$\mathbf{p}$	Optical momentum: $\mathbf{p} = n\mathbf{t}$ where n is the refractive index and $\mathbf{t}$ is a unit vector tangent to the light ray. It has components $\mathbf{p} = (p_1, p_2, p_3)$.
$\mathbf{s}(\sigma)$	Path of a light ray with parameter σ
S	Optical path length
u_1, u_2, u_3	Generalized momenta
U	Étendue
$\mathbf{\Phi}$	Flux (energy per unit time)
Ω	Solid angle
P	Hamiltonian (when light paths are parameterized by a generic parameter σ)

P	Point **P**. It has coordinates $\mathbf{P} = (P_1, P_2, P_3)$ along the axes x_1, x_2, x_3 axes
$\mathbf{P}_1$	Point $\mathbf{P}_1$. It has coordinates $\mathbf{P}_1 = (P_{11}, P_{12}, P_{13})$ along the axes x_1, x_2, x_3
v	Vector $\mathbf{v} = (v_1, v_2, v_3)$
x_1, x_2, x_3	Spacial coordinates
//	Parallel to (in some figures)

List of Abbreviations

2-D	Two-dimensional geometry
3-D	Three-dimensional geometry
CAP	Concentration Acceptance Product
CPC	Compound Parabolic Concentrator
CEC	Compound Elliptical Concentrator
DTIRC	Dielectric Total Internal Reflection Concentrator
I	Total Internal Reflection in the SMS design method
R	Refraction in the SMS design method
RR	SMS lens made of two refractive surfaces
RX	SMS optic made of a refractive and a reflective surface
RXI	SMS optic in which light undergoes a refraction, then a reflection, and then a Total Internal Reflection (TIR)
SMS	Simultaneous Multiple Surfaces (design method)
TERC	Tailored Edge Ray Concentrator
TIR	Total Internal Reflection
X	Reflection in the SMS design method
XR	SMS optic made of a reflective and a refractive surface
XX	SMS optic made of two reflective surfaces

Angle Transformer	Device that accepts radiation with a given angle θ_1 and puts out radiation with another angle θ_2
Angle Rotator	Device that rotates light, maintains the area and angle of the light
Trumpet	Concentrator made of two hyperbolic mirrors with the same foci. All the radiation headed toward the line between the foci and intersected by the mirrors is concentrated onto the line between the vertex of the hyperbolas

Functions defined in Chapter 21:
nrm(…), ang(…), angp(…), angpn(…), angh(…), $R(\alpha)$, islp(…), isl(…), par(…), eli(…), hyp(…), winv(…), uinv(…), wmp(…), ump(…), wme(…), ume(…), cop(…), cco(…), dco(…), coptpt(…), ccoptpt(…), dcoptpt(…), coptsl(…), rfx(…), rfr(…), rfrnrm(…), rfxnrm(…).

Section I

Nonimaging Optics

1

Why Use Nonimaging Optics

1.1 Area and Angle

Typically, optical systems have an emitter (light source) *E* which emits light, an optic *O* which deflects it, and a receiver (target) *R* which is illuminated by it, as shown in Figure 1.1.

As light travels in space, or through an optic, it spreads over some area, and also over some angle. Area and angle are related and that is the basis for illumination optics. Figure 1.2 shows some flashlights emitting light downwards. In Figure 1.2a, the flashlights emit light toward a large area a_1. The angular aperture θ_1 of the light crossing a_1 is small. In Figure 1.2b, the same flashlights now emit light toward a small area a_2. The angular aperture θ_2 of the light crossing a_2 is now large. The reason for this is that we cannot place all flashlights on top of one another; they must be placed side by side, and this increases the angle of the light going through a_2.

If light crosses a large area contained within a small angular aperture (as in Figure 1.2a), and we now try to "squeeze" that light through a small area (as in Figure 1.2b), its angular aperture increases. So, for the same amount of light crossing an aperture, if the area is large, the angle is small; and if the area is small, the angle is large. This characteristic of light is called conservation of étendue.

As referred above, in an illumination system, we have an emitter (source), an optic, and a receiver (target). The emitter and receiver may be of different sizes. If the emitter *E* is small and the receiver *R* is large and distant, the optic is called a collimator, as in Figure 1.3a. If the emitter *E* is large and distant and the receiver *R* is small, the optic is called a concentrator, as in Figure 1.3c. When emitter *E* and receiver *R* are of similar size, the optic is a condenser, as in Figure 1.3b.

An example of a collimator is a flashlight emitting light rays *r* coming from a small source inside it and illuminating, for example, a large wall *R*, as in Figure 1.4a. An example of a concentrator is a magnifying glass pointed at the sun (the emitter), collecting sun rays *r* and concentrating sunlight onto a small receiver *R*, as in Figure 1.4b.

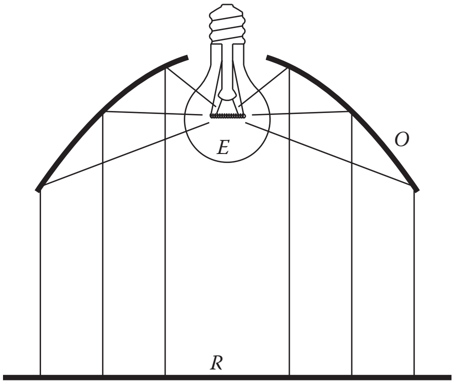


FIGURE 1.1
Optical system composed of an emitter *E*, an optic *O*, and a receiver *R*.

Figure 1.5a shows an optic (lens in this example) with emitter *E* and receiver *R*. Figure 1.5b shows a diagrammatic representation of the same. For the sake of simplicity, this diagrammatic representation will be used below.

Two different optics, O_1 and O_2, may capture the same amount of light from an emitter *E*, as shown diagrammatically in Figure 1.6. These two optics capture light from the same size emitter *E* within the same emission angle α.

Different size optics for the same purpose may have an important impact on the system in which these optics are included.

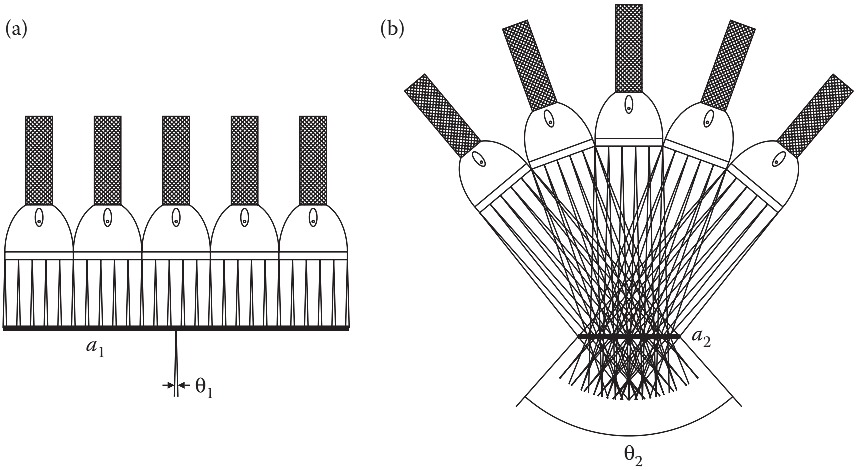


FIGURE 1.2
The same amount of light emitted by some flashlights either (a) crosses a large area a_1 with a small angle θ_1 or (b) crosses a small area a_2 with a large angle θ_2.

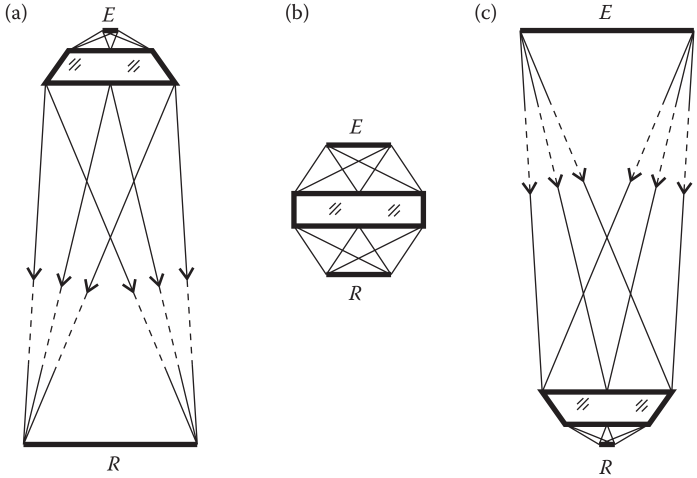


FIGURE 1.3
(a) If the emitter *E* is small and the receiver *R* is large, the optic is a collimator. (b) If emitter *E* and receiver *R* are of similar size, the optic is a condenser. (c) If the emitter *E* is large and the receiver *R* is small, the optic is a concentrator.

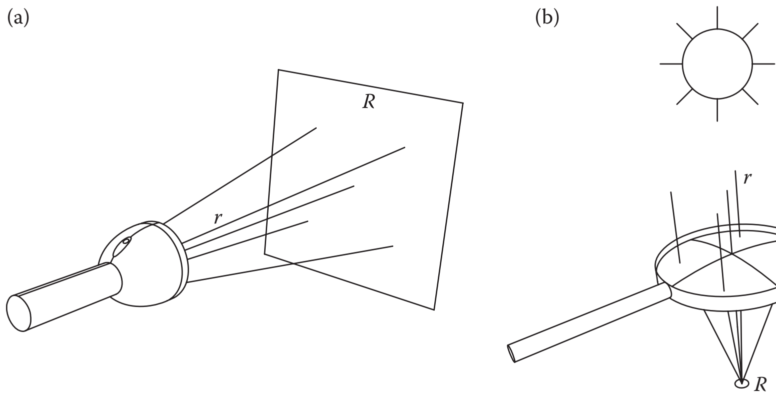


FIGURE 1.4
(a) Flashlight, example of a collimator. (b) Magnifying glass pointing at the sun, example of a concentrator.

1.2 Collimators: Illumination of a Large Receiver

Consider the situation shown in Figure 1.7a with a small emitter *E* emitting within angle α and a large receiver *R*. The light from *E* can be transferred to *R* by a collimator optic O_1, as shown in Figure 1.7b. However, it may also be that a smaller collimator optic O_2 can also transfer the same light from *E* to *R*.

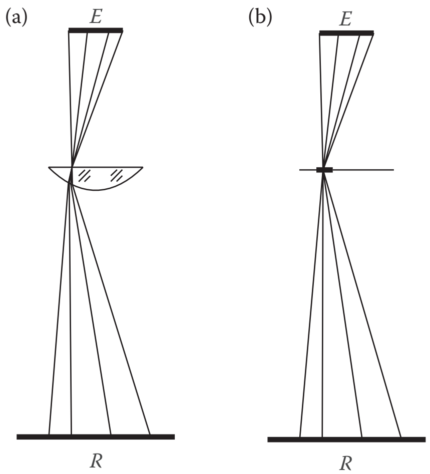


FIGURE 1.5
(a) Optic (lens) with emitter E and receiver R. (b) Diagrammatic representation of the same.

From a practical point of view, however, a smaller collimator optic may be preferable, since it may be cheaper to make, package, or ship, and needs less volume to be installed into.

Typically, it is advantageous to have collimator optics which are efficient and small, since that minimizes cost. By efficient, we mean an optic which transfers to the receiver all the light it intersects from the emitter.

As seen above in Figure 1.2, if we try to "squeeze" light through a small area, its angular aperture increases. This is called conservation of étendue.

Figure 1.8a shows a small emitter E, a large receiver R, and a collimator optic O_1. Now consider a small section da_1 of O_1. Light crossing da_1 is contained between rays r_1 and r_2, coming from the edges of emitter E, and leaves da_1 contained within angle θ_1. This optic is efficient, since it redirects to R all the light it intersects from E.

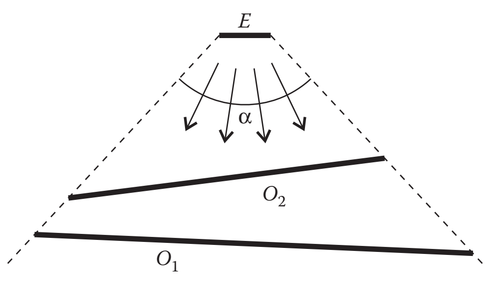


FIGURE 1.6
Optics O_1 and O_2 "see" the same size emitter E and same emission angle α and, therefore, capture the same amount of light from E.

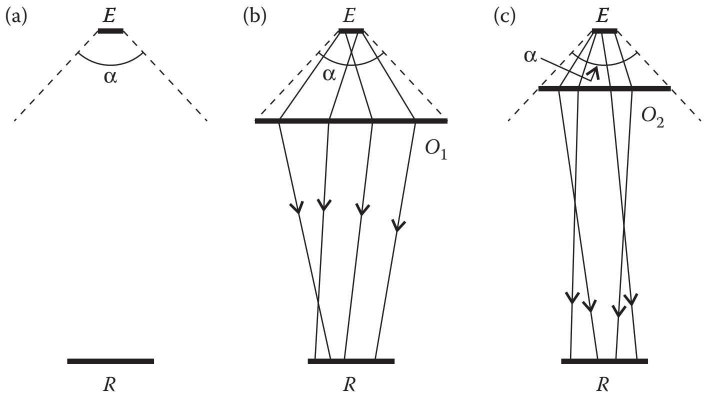


FIGURE 1.7
(a) Small emitter E emitting light within an angle α and a large receiver R. (b) Large collimator optic O_1 captures the light from E and transfers it to receiver R. (c) Small collimator optic O_2 captures the light from E and transfers it to R.

Now consider a smaller optic O_2 capturing the same amount of light from E and redirecting it to receiver R, as shown in Figure 1.8b. Since now O_2 is smaller than O_1, a section da_2 of O_2 is also smaller than the corresponding section da_1 of O_1. The amount of light from E intersected by da_2, is the same as the amount of light from E intersected by da_1. However, due to the conservation of étendue, if area da_2 is smaller than da_1, angle θ_2 is larger than θ_1. In this case, rays r_1 and r_2, coming from the edges of emitter E, are redirected toward the edges of receiver R. This optic is efficient, since it redirects to R all the light it intersects from E.

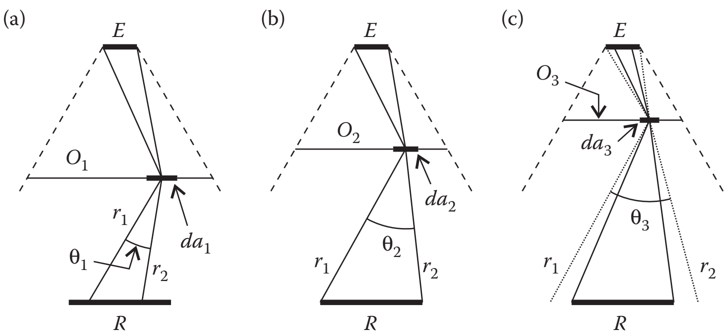


FIGURE 1.8
(a) Light from E crossing a large section da_1 of large optic O_1 falls inside receiver R. (b) Optimum situation in which the rays from the edges of E are redirected toward the edges of R. (c) Inefficient optic since some light from E crossing a small section da_3 of small optic O_3 fails receiver R.

Finally, consider yet another smaller optic, O_3, capturing the same amount of light from *E*, and redirecting it to receiver *R*, as shown in Figure 1.8c. Since now O_3 is smaller than O_2, a section da_3 of O_3 is also smaller than the corresponding section da_2 of O_2. However, due to the conservation of étendue, if area da_3 is smaller than da_2, then angle θ_3 is larger than θ_2. Since, in Figure 1.8b, rays r_1 and r_2 were redirected to the edges of receiver *R*, and θ_3 in Figure 1.8c is now larger than θ_2, these rays r_1 and r_2 must be redirected by O_3 in wider directions that fail receiver *R*. This optic is inefficient, since it fails to redirect to *R* all the light it intersects from *E*.

The optimum solution is, therefore, that given in Figure 1.8b. At da_2, the maximum angle possible without losing light, is θ_2, in which rays r_1 and r_2 (coming from the edges of emitter *E*) are redirected to the edges of receiver *R*. This is called the edge ray principle. A smaller angle θ_1 still puts all light on *R*, and results in an efficient optic, but da_1 (and optic O_1) could be smaller. On the other hand, a wider angle θ_3 results in a small da_3 (and small optic O_3), but some light fails *R*, resulting in a low efficiency optic.

It may then be concluded that, for collimators, efficiency and compactness combined result in the edge ray principle.

Figure 1.9 shows the same optical configuration as Figure 1.8, only now with real lenses, instead of a diagrammatic representation.

Section da_2 of the optic in Figure 1.8b redirects to the edges of the receiver *R* the rays coming from the edges of the emitter *E*. Since da_2 may be at any position across the optic, an optic with these characteristics is as shown in Figure 1.10a. This optic obeys the edge ray principle: rays coming from edge $\mathbf{E}_1$ of emitter *E* and crossing any position $\mathbf{P}$ on the optic are redirected to edge

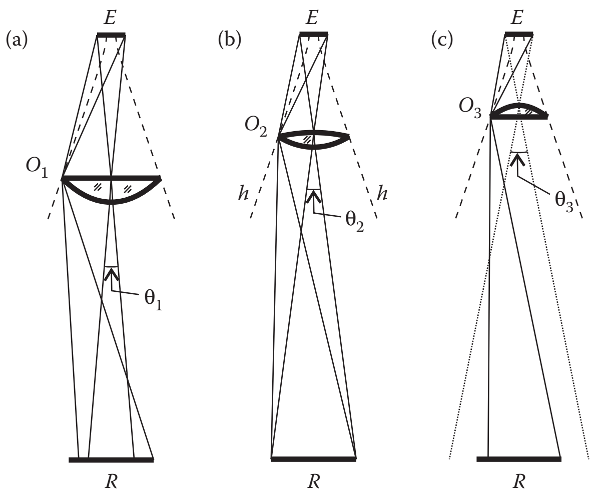


FIGURE 1.9
Same as Figure 1.8, only now with real lenses instead of a diagrammatic representation. (a) Efficient but large and optic. (b) Optimum optic. (c) Small but inefficient optic.

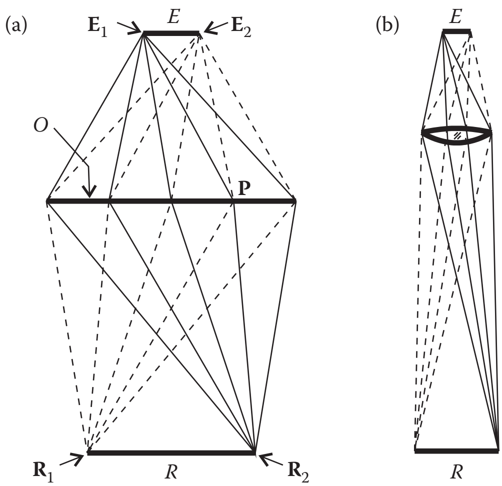


FIGURE 1.10
(a) Diagrammatic representation of a collimator optic O that redirects the rays coming from the edges of the emitter E to the edges of the receiver R. (b) Real lens instead of a diagrammatic representation.

$\mathbf{R}_2$ of the receiver R, and, accordingly, rays coming from edge $\mathbf{E}_2$ of emitter E and crossing any position $\mathbf{P}$ on the optic, are redirected to edge $\mathbf{R}_1$ of the receiver R. Figure 1.10b shows the same, but with a real lens instead of a diagrammatic representation.

One option to transfer light from an emitter to a receiver is to focus the center of the emitter E onto the center of the receiver R, like optic O_1 in Figure 1.11a. In general, however, this optic O_1 will not satisfy the edge ray principle, and the rays coming from the edges of the emitter E are not focused at the edges of the receiver R, as shown in Figure 1.11b, in which the rays coming from the edges of E spread over R. This optic is large and must be placed far away from the emitter. For that reason, emitter E and optic O_1 use a large volume V_1. Optic O_1 is designed for a point source (emitter) at the center of the extended emitter E. Also, O_1 is large when compared to emitter E. For those reasons, O_1 is called a point source optic.

However, it is possible to design a different optic O_2, which verifies the edge ray principle as shown in Figure 1.11c. Optic O_2 captures the same amount of light from E as O_1 and transfers it to receiver R (as does O_1). However, optic O_2 and emitter E now use a smaller volume V_2. Also, O_2 is designed for the edges of source (emitter) E and, therefore, its full extent. For that reason, O_2 is called an extended source optic.

The receiver R of a collimator O_1 may have a very large size, and be at a very large distance from O_1, as shown in Figure 1.12. In the limit case we get an infinite receiver R placed at an infinite distance. An extended source collimator optic (Figure 1.12a) must emit light confined to an angle α to fully

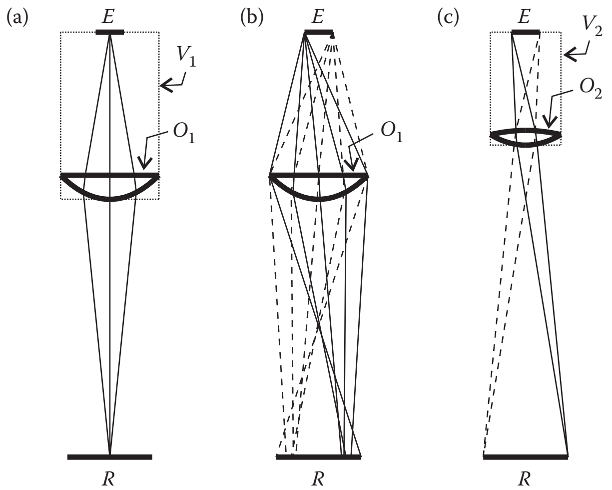


FIGURE 1.11
(a) Optic O_1 is designed to focus the center of E onto the center of R; it needs a large volume V_1. (b) Optic O_1 does not obey the edge ray principle and rays from the edges of E do not converge to the edges of R. (c) Optic O_2 designed with the edge ray principle needs a small volume V_2.

illuminate the receiver. Receiver R subtends an angle α when "seen" from optic O_1. Light emitted from O_1 at a wider angle than α will fail receiver R, resulting in low efficiency.

A point source collimator O_2 (Figure 1.12b) focuses the center of the source (emitter) E onto the center of the target (receiver) R. In the limit case in

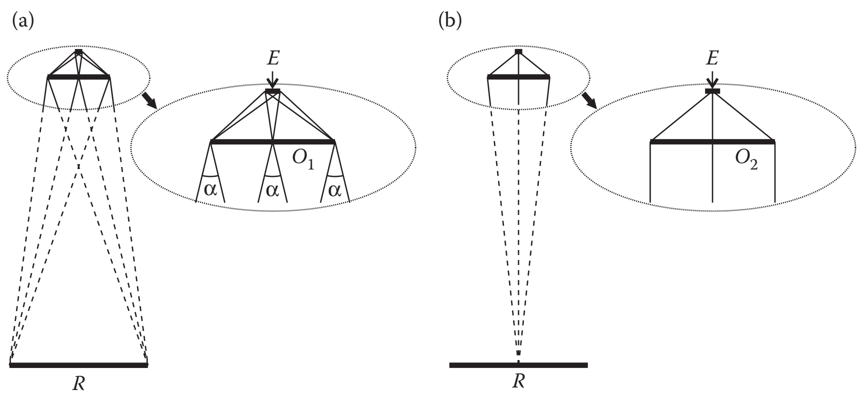


FIGURE 1.12
Receiver R is very large and very far away from optics O_1 and O_2. In the limited case, we have an infinite size receiver R placed at an infinite distance. (a) Extended source optic O_1 must emit light within an angle α to illuminate R. (b) Point source optic O_2 takes the rays from the center of E and emits them parallel to each other toward the center of R.

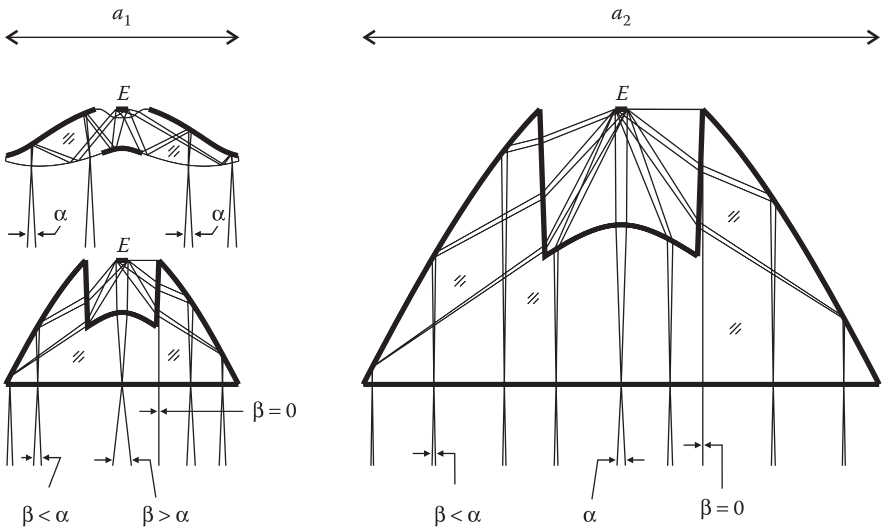


FIGURE 1.13
Left, top: optic designed with the edge ray principle, compact and efficient. Left, bottom: point source optic, compact but inefficient. Right: point source optic, efficient but very large.

which R is at an infinite distance from O_2, the rays coming from the center of E are emitted parallel to each other.

As an example of different kinds of collimator optics, Figure 1.13 shows three optics for the same emitter E and a large and distant (infinite) receiver R that subtends an angle α when "seen" from these optics.

The optic on the left, top, is an extended source optic (as in Figure 1.12a) designed with the edge ray principle. Its exit aperture is a_1, and its angular emission angle is α across its whole aperture. This optic, called an RXI[1] (see Chapter 9), is compact and efficient. The optic on the left, bottom, is a point source optic (as in Figure 1.12b) with the same emitter E and same exit aperture a_1. At some positions, it emits light within a cone $\beta < \alpha$, and therefore, all that light will fall on the receiver. However, at other positions, it emits light within a cone $\beta > \alpha$, and therefore, some of that light will fail the receiver and be lost. This optic is therefore compact but inefficient. The optic on the right is again a point source optic (as in Figure 1.12b) with the same emitter E but with a much larger exit aperture a_2. Its emission angle is $\beta \leq \alpha$ across the whole aperture a_2. All the light will then hit the receiver. This optic is therefore efficient but it is not compact.

Figure 1.14 shows an RXI (left) and a point source optic (right), both with the same exit aperture a_1. It illustrates the comparison in Figure 1.13 on the left.

Note that in Figure 1.14 the optics point up, while in Figure 1.13 the optics point down.

FIGURE 1.14
RXI (left) and point source optic (right), both with the same diameter. (Courtesy of Light Prescriptions Innovators.)

1.3 Concentrators: Illumination of a Small Receiver

Small collimator optics are desirable, but that is not the case with concentrator optics. Actually, the opposite is true. In a concentrator, the larger its aperture, the more light it captures from the source (emitter). Figure 1.15a shows diagrammatically two concentrator optics, O_1 and O_2, for large emitter E and small receiver R. Both optics "see" the same size receiver R at the same angle α and, therefore, have the same "ability" to put light onto R. Optic O_2, however, is larger and, therefore, captures more light from the emitter E.

Figure 1.15b shows the same, but for solar concentrators O_1 and O_2. By being larger, optic O_2 captures more sunlight and redirects it to receiver R.

Typically, it is advantageous to have concentrator optics that are efficient and large, since that maximizes the amount of light captured from the emitter. By "efficient," we mean an optic which transfers to the receiver all the light it intersects from the emitter.

Figure 1.16a shows a large emitter E, a small receiver R, and a concentrator optic O_1. Now consider a small section da_1 of O_1. Light crossing da_1 contained between rays r_1 and r_2 would fall on receiver R. Since O_1 is small, da_1 is also small, and therefore, angle θ_1 is large (conservation of étendue). Rays r_1 and r_2 would fall on R, but the emitter E does not extend that far and, therefore, no light is coming from those directions. However, this optic is efficient, since it redirects to R all the light it intersects from E.

Now consider a larger optic O_2, as shown in Figure 1.16b. Since now O_2 is larger than O_1, a section da_2 of O_2 is also larger than the corresponding section da_1 of O_1. However, due to the conservation of étendue, if area da_2 is larger than da_1, angle θ_2 is smaller than θ_1. In this case, rays r_1 and r_2 coming

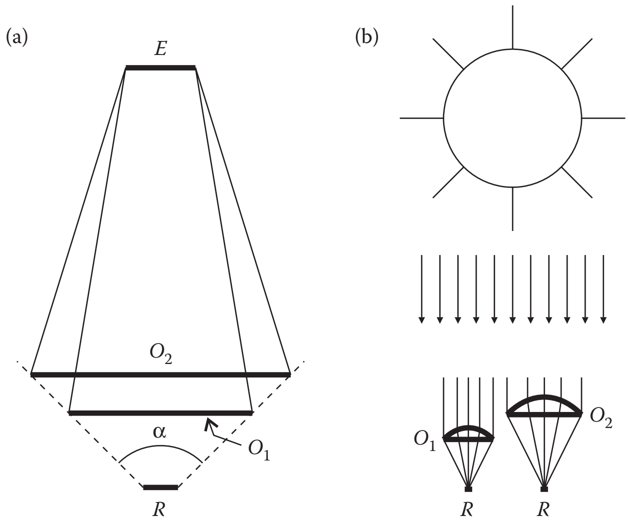


FIGURE 1.15
(a) A larger concentrator O_2 captures and redirects to receiver *R* more light from the emitter *E* than a smaller optic O_1. (b) Situation in which optics O_1 and O_2 are solar concentrators and the emitter is the sun.

from the edges of emitter *E* are redirected toward the edges of receiver *R*. This optic is efficient, since it redirects to *R* all the light it intersects from *E*.

Here, the "ability" to transfer light from da_1 to *R* is the same as that from da_2 to *R*. The étendue of the light transferable between da_1 and *R* is the same as the étendue of the light transferable between da_2 and *R*. However, in the case of Figure 1.16a, that "ability" is not fully used, since no light is coming from directions r_1 or r_2.

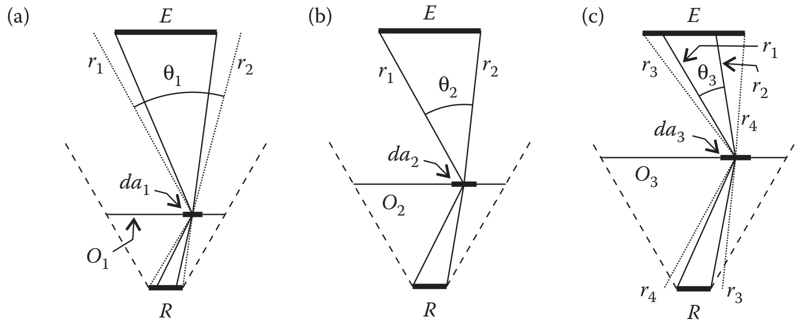


FIGURE 1.16
(a) A small section da_1 of small optic O_1 could accept light from directions r_1 and r_2 beyond emitter *E* and where there is no light. (b) Optimum situation in which the rays from the edges of *E* are redirected toward the edges of *R*. (c) Inefficient optic since some light from *E* crossing a large section da_3 of large optic O_3 fails receiver *R*.

Finally, consider yet another larger optic O_3 as shown in Figure 1.16c. Since now O_3 is larger than O_2, a section, da_3, of O_3 is also larger than the corresponding section da_2 of O_2. However, due to the conservation of étendue, if area da_3 is larger than da_2, angle θ_3 is smaller than θ_2. In Figure 1.16b, rays r_1 and r_2 reaching the edges of receiver R were coming from the edges of the emitter E. Since angle θ_3 is now smaller, these rays must come from points inside the emitter E. However, now rays r_3 and r_4 coming from the edges of emitter E and crossing da_3 will fail the receiver R. This optic is then inefficient, since it fails to redirect to R all the light it intersects from E.

The optimum solution is, therefore, that given in Figure 1.16b. At da_2, the minimum angle possible, without losing light, is θ_2, in which rays r_1 and r_2 (coming from the edges of emitter E) are redirected to the edges of receiver R. This is again the edge ray principle. A larger angle θ_1 still puts all light on R and results in an efficient optic, but da_1 (and optic O_1) could be larger, capturing more light from E. On the other hand, a smaller angle, θ_3, results in a larger da_3 (and larger optic O_3) but some light fails R, resulting in a low-efficiency optic.

It may then be concluded that, for concentrators, efficiency and maximum size, combined, result in the edge ray principle.

Figure 1.17 shows the same as Figure 1.16, only now with real lenses instead of a diagrammatic representation.

Section da_2 of the optic in Figure 1.16b redirects to the edges of the receiver R the rays coming from the edges of the emitter E. Since da_2 may be at any position across the optic, an optic with these characteristics is as shown in Figure 1.18a. This optic obeys the edge ray principle: rays coming from edge $\mathbf{E}_1$ of emitter E and crossing any position $\mathbf{P}$ on the optic are redirected to edge

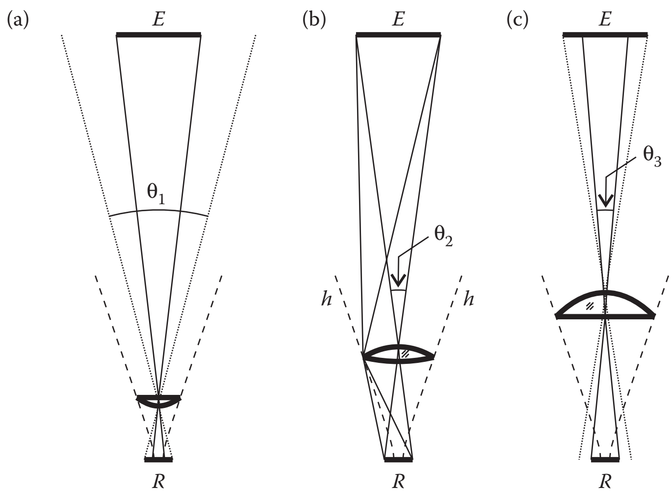


FIGURE 1.17
Same as Figure 1.16 only now with real lenses instead of a diagrammatic representation. (a) Efficient but small optic. (b) Optimum optic. (c) Large but inefficient optic.

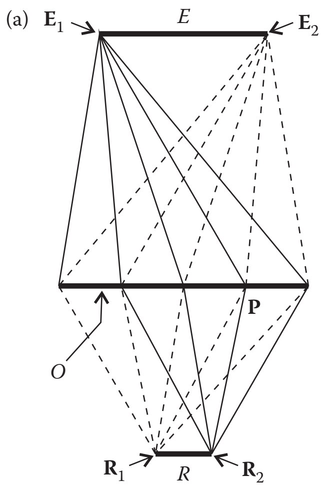


FIGURE 1.18
(a) Diagrammatic representation of a concentrator optic O that redirects the rays coming from the edges of the emitter E to the edges of the receiver R. (b) Real lens instead of a diagrammatic representation.

$\mathbf{R}_2$ of the receiver R, and accordingly rays coming from edge $\mathbf{E}_2$ of emitter E and crossing any position **P** on the optic are redirected to edge $\mathbf{R}_1$ of the receiver R. Figure 1.18b shows the same optical configuration, but with a real lens, instead of a diagrammatic representation.

The emitter E of a concentrator, O_1, may have a very large size, and be at a very large distance from O_1, as shown in Figure 1.19. In the limit case of an

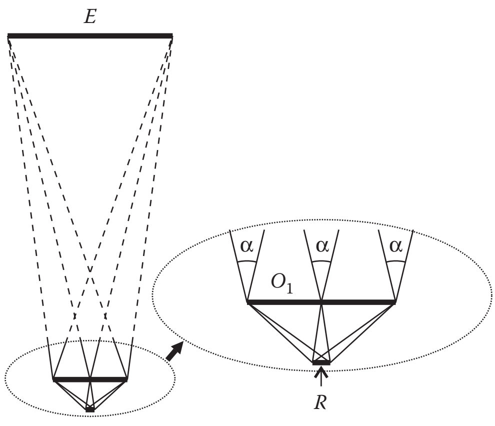


FIGURE 1.19
Emitter E is very large and very far away from concentrator optic O_1. In the limited case of an infinite size emitter at an infinite distance, the incoming rays make an angle α to each other across the whole aperture of O_1. Angle α is called total acceptance angle.

infinite size emitter at an infinite distance, emitter E has an angular aperture α, that is, it subtends an angle α when "seen" from optic O_1. Light rays coming from the edges of E make an angle α to each other when they reach O_1. Light contained inside angle α is redirected by the concentrator to its receiver R. Angle α is the total acceptance angle of O_1.

As an example of different kinds of concentrator optics, Figure 1.20 shows two optics for the same circular receiver R and same emitter angular aperture α.

The optic composed of a primary mirror m_1 and secondary mirror m_2 is designed with the edge ray principle.[2] It has a constant acceptance angle α across its large aperture a_1.

The parabolic mirror p has an acceptance angle α at the edge, but a wider acceptance angle β at other points. It still redirects all the light, it intersects from the emitter, to the receiver (which has angular aperture $\alpha < \beta$), but this increased acceptance angle β results in a small aperture a_2. Edge rays for circular receivers R will be discussed in other chapters.

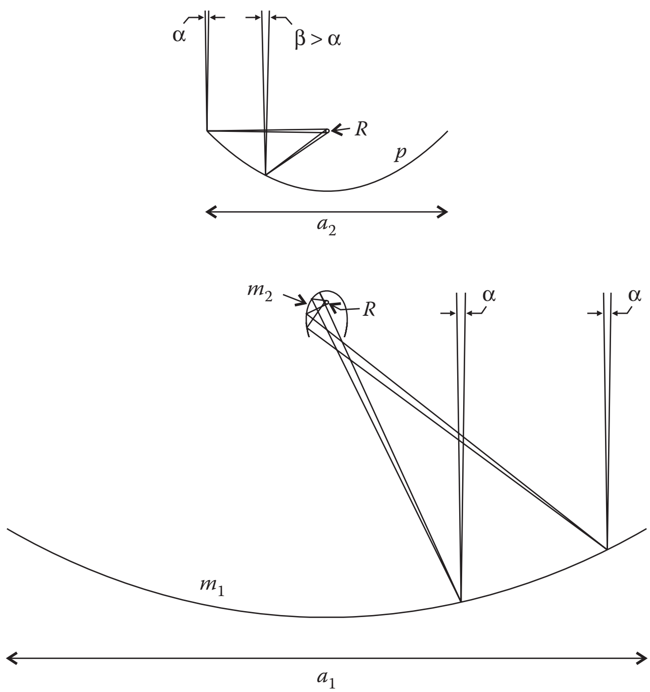


FIGURE 1.20
Large optic made of a primary mirror m_1 and secondary mirror m_2 has a large aperture a_1 and the same acceptance angle α across the whole aperture. Small parabolic mirror p has the same size receiver R but a smaller aperture a_2.

1.4 Collimators and Concentrators Summary

The results obtained above for collimators and concentrators may now be summarized in Figures 1.21 and 1.22. In the case of collimators, as shown in Figure 1.21, the optimum optical solution is the one in the middle, designed with the edge ray method. To the left of it, we have optics which are efficient but large, and to the right of it we have optics which are small but inefficient. All these optics capture the same amount of light from emitter *E*.

In the case of concentrators, as shown in Figure 1.22, the optimum optical solution is again the one in the middle, designed with the edge ray method. To the left of it, we have optics which are large but inefficient, and to the right of it we have optics which are efficient but small. All these optics have the same "ability" to put light onto receiver *R*.

An optic designed with the edge ray principle focuses the light from edge $\mathbf{E}_1$ of the source onto the edge $\mathbf{R}_2$ of the receiver, and the light from edge $\mathbf{E}_2$ of

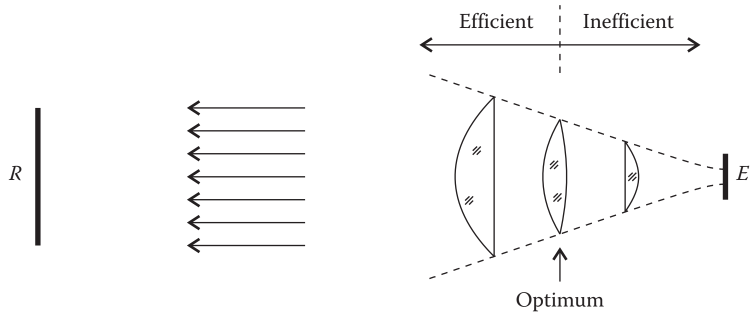


FIGURE 1.21
Collimator optics with emitter *E* and receiver *R*. Optimum optic at the center. To the left, optics are efficient, but large. To the right, optics are small but inefficient.

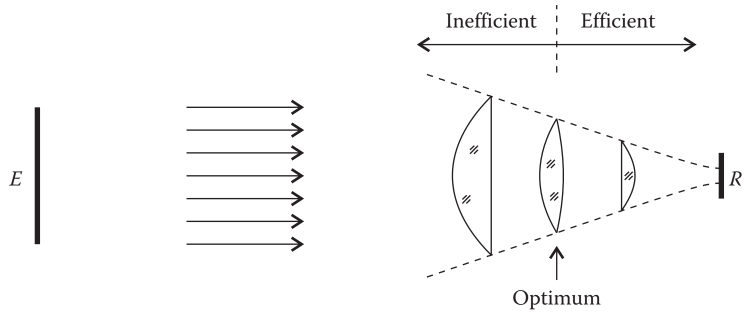


FIGURE 1.22
Concentrator optics with emitter *E* and receiver *R*. Optimum optic at the center. To the left, optics are large but inefficient. To the right, optics are efficient but small.

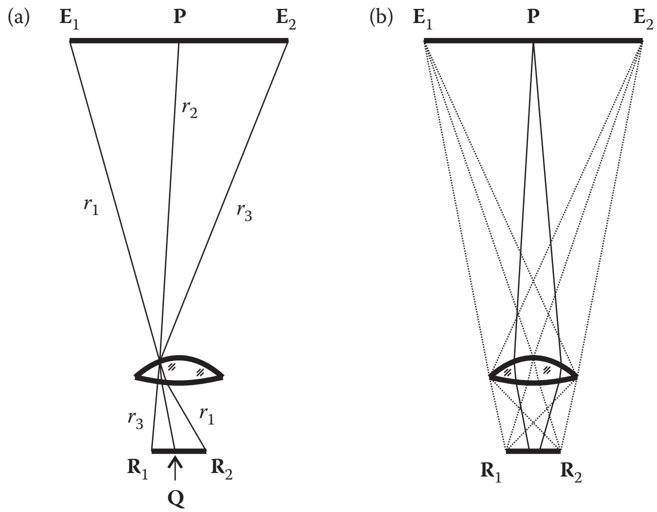


FIGURE 1.23
(a) Rays r_1 and r_3 coming from the edges of the emitter are redirected to the edges of the receiver and, therefore, a ray r_2 coming from inside the emitter is redirected to a point inside the receiver. (b) An optic designed with the edge ray principle does not necessarily focus a point **P** in the emitter onto a point **Q** in the receiver. There is no need for image formation and the optic is called nonimaging.

the source onto the edge $\mathbf{R}_1$ of the receiver, as shown in Figure 1.23. However, there is no condition for the light emitted from a point **P** inside the receiver. The light emitted from a point **P** in the emitter does not need to converge on a point at the receiver and, therefore, there is no condition for image formation, as shown in Figure 1.23b. For that reason, these are called nonimaging optics.

However, if rays r_1 and r_3 coming from the edges of the emitter are redirected to the edges of the receiver, a ray r_2 coming from inside the emitter will end up inside the receiver, as shown in Figure 1.23a. For that reason, the edge ray principle of nonimaging optics ensures that all light from the emitter that is intersected by the optic is transferred to the receiver.

1.5 Collimators Tolerances

Optics are first designed and then manufactured. Manufacturing, however, introduces different kinds of errors, which include imperfections in the optical surfaces, or imperfect assembly of components. Usage also introduces further errors, such as relative movements of the components or deterioration over time. There are, therefore, many sources of errors when making and using optics. It is, therefore, highly desirable to design the optics which are as tolerant as possible to all these errors.[3]

High tolerance to errors leads to relaxed manufacture, assembly and usage, which in turn leads to lower cost.

Figure 1.24a shows a perfect optic which focuses a point in the emitter *E* onto a point on the receiver *R*. If this optic is made, the manufactured optic may have errors—for example, oscillations on optical surface *s*, as shown in Figure 1.24b. Light is no longer focused at a point on *R*, but, instead, spreads out over some area. Errors are quite often random variations, and light is deviated randomly (diffused) relative to its ideal path.

Now consider a nonimaging collimator with emitter *E*, receiver *R*, and fixed aperture size **AB**, as shown in Figure 1.25a. This optic fully illuminates the receiver *R* since the rays coming from the edges of *E* are redirected to the edges of *R*. Suppose now that we wanted to decrease the spot size on *R*, that is, we wanted to put all the light in a smaller spot inside of *R*. One could then increase the area of a small section of the collimator from a_1 to a_3. From conservation of étendue, angle α_1 would decrease to α_3, and a_3 would illuminate a smaller spot inside *R*, as desired (see Figure 1.25b). However, since the total aperture **AB** is fixed, somewhere else in the optic, another area a_2 must decrease to an area a_4. Again from conservation of étendue, angle α_2 must increase to α_4, and this will make the light to spread over a larger area, increasing the spot to a larger size *s*. It can then be concluded that the spot created by a nonimaging optic

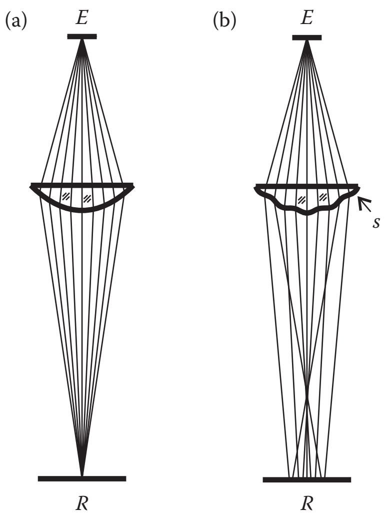


FIGURE 1.24
(a) Perfect optic focuses a point of emitter *E* onto a point of receiver *R*. (b) Imperfectly manufactured optic: oscillations in one of its surfaces *s* scatter the light as it passes the optic, which no longer focuses the light onto a point.

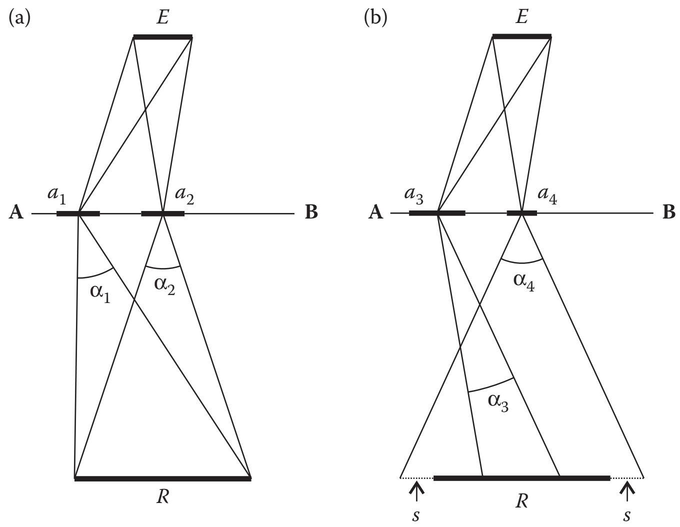


FIGURE 1.25
(a) Nonimaging optic collimator of fixed aperture size **AB**. (b) By increasing a_1 to a_3 then α_1 decreases to α_3 illuminating a smaller spot in *R*. However a_2 decreases to a_4 and α_2 increases to α_4 leading to larger spot size *s*.

cannot be decreased. The nonimaging optic in Figure 1.25a then minimizes the angular aperture α across its whole aperture.

Figure 1.26 shows the same as Figure 1.25, only now with real lenses, instead of a diagrammatic representation.

We have seen that errors (in manufacture, assembly, or usage) diffuse the light as it crosses the optic, therefore increasing the spot size. However, with nonimaging optics, we start with the minimum spot size. Therefore, we also end up with the minimum spot size when these errors are introduced. It may then be concluded that nonimaging optics maximize tolerances to errors in collimators.

Figure 1.27a shows an ideal nonimaging optic with emitter *E* and aperture **AB** that creates a spot inside receiver *R*. It is designed for a smaller receiver than *R*. Figure 1.27b shows the same optic, but now manufactured, assembled, and used. Errors in these processes now diffuse the light that now fully illuminates the whole receiver *R*.

Figure 1.27c shows a nonideal optic with emitter *E* and aperture **AB** that also creates a spot inside the receiver *R*. However, since this optic was not designed with the edge ray principle, the spot it creates is now larger than the spot created by the optic in Figure 1.27a. Figure 1.27d shows the same optic in Figure 1.27c, but now as a real optic. Errors now diffuse the light that spreads over an area larger than the receiver *R*, as shown by rays *r*. Since some light now fails to reach the receiver, efficiency decreases.

It may then be concluded that the nonimaging optic in Figure 1.27a is more tolerant to errors, and maintains efficiency. Even with errors, all light falls on

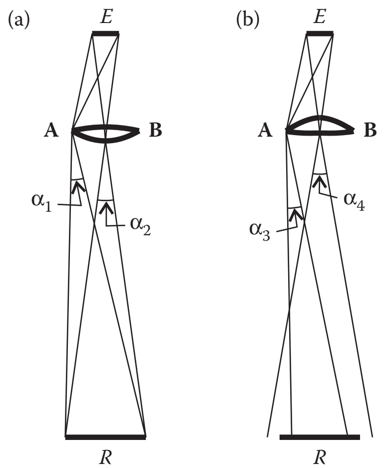


FIGURE 1.26
Same as Figure 1.25, only now with real lenses instead of a diagrammatic representation. (a) Nonimaging optic. (b) Optic with reduced angle α_3 at some position results in an increased angle α_4 somewhere else if the aperture **AB** is fixed.

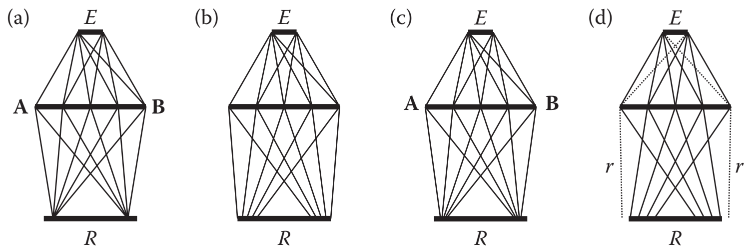


FIGURE 1.27
(a) Ideal nonimaging optic creates an ideal (minimum size) spot inside the receiver *R*. (b) Errors in the manufactured nonimaging optic diffuse the light, which now illuminates the whole receiver *R*. (c) Nonideal optic creates a larger than ideal spot on receiver *R*. (d) Errors in the manufactured nonideal optic diffuse the light which now spills out of *R* resulting in light loss.

the receiver. However, the nonideal optic in Figure 1.27c is not as tolerant to errors, and does not maintain efficiency. With errors, some light will miss the receiver, lowering efficiency.

1.6 Concentrators Tolerances

Concentrators, like collimators, will have imperfections when they are made, assembled, and used. Also in the case of concentrators, errors will diffuse

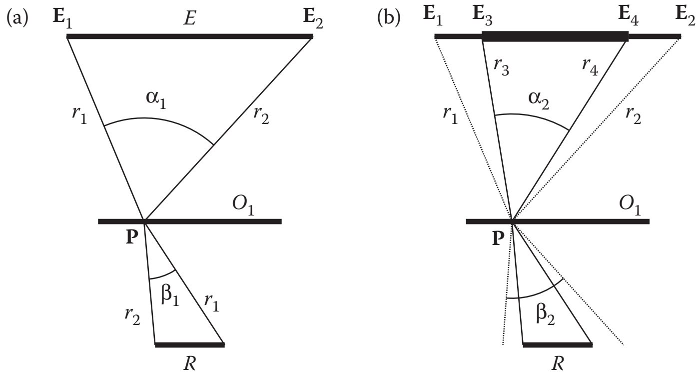


FIGURE 1.28
(a) At position **P** of optic O_1 the acceptance angle is α_1 and light leaves **P** within angle β_1. (b) Errors diffuse the light, which now leaves **P** within a wider angle β_2. Acceptance angle is reduced to α_2.

the light as it passes the optic. Figure 1.28a shows a nonimaging concentrator O_1. Incoming rays inside acceptance angle α_1 at position **P** leave inside angle β_1, fully illuminating the receiver R. When this optic is made and used, errors will diffuse the light as it passes the real optic, as shown in Figure 1.28b. Now, the incoming light contained between rays r_1 and r_2 at position **P** will leave the optic contained in a wider angle β_2, and some of that light will miss the receiver R, lowering efficiency.

The light crossing position **P** that does reach the receiver is now contained between rays r_3 and r_4, coming from a smaller area $\mathbf{E}_3\mathbf{E}_4$ inside emitter E (which extends from $\mathbf{E}_1$ to $\mathbf{E}_2$). Errors then result in a reduced acceptance angle α_2 at position **P**.

Now consider a nonimaging concentrator with emitter E, receiver R, and fixed aperture size **AB**, as shown in Figure 1.29a. This optic captures light from the whole emitter E, since the rays coming from the edges of E are redirected to the edges of R. Suppose now that we wanted to increase the acceptance angle α so that the optic could capture light from an emitter larger than E. One could then decrease the area of a small section of the collimator from a_1 to a_3. From conservation of étendue, angle α_1 would increase to α_3, and a_3 would have a wider acceptance angle α_3, as desired (see Figure 1.29b). However, since the total aperture **AB** is fixed, somewhere else in the optic, another area a_2 must increase to an area a_4. Again from conservation of étendue, angle α_2 must decrease to α_4, and this will reduce the acceptance angle at a_4. Now some rays, including r_1 and r_2, coming from the outer portions of emitter E will miss the receiver, reducing efficiency. It can then be concluded that the acceptance angle of a nonimaging optic cannot be increased. The nonimaging optic in Figure 1.29a then maximizes the acceptance angle α across its whole aperture.

Figure 1.30 shows the same optical configuration as Figure 1.29, only now with real lenses instead of a diagrammatic representation.

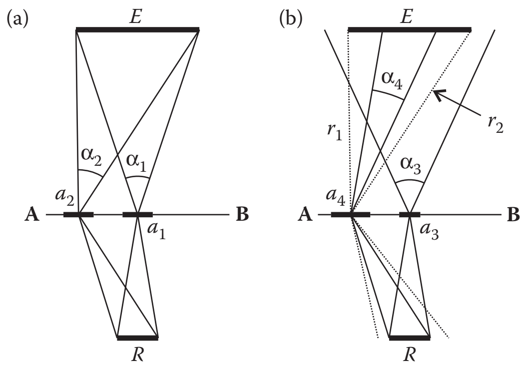


FIGURE 1.29
(a) Nonimaging optic concentrator of fixed aperture size **AB**. (b) By decreasing a_1 to a_3 then α_1 increases to α_3 increasing the acceptance angle. However a_2 increases to a_4 and α_2 decreases to α_4 leading to light loss of rays including r_1 and r_2.

We have seen that errors (in manufacture, assembly, or usage) diffuse the light as it crosses the optic, thereby reducing the acceptance angle. However, with nonimaging optics, we start with the maximum acceptance angle. Therefore, we also end up with the maximum acceptance angle when these errors are introduced. It may then be concluded that nonimaging optics maximize tolerances to errors in concentrators.

Figure 1.31a shows a nonimaging concentrator optic with aperture **AB** and acceptance angle α_1 at position **P**, wider than the angular aperture of emitter *E*. When this optic is manufactured and used, errors diffuse the light and

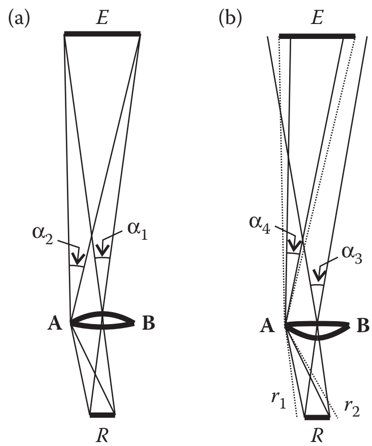


FIGURE 1.30
Same as Figure 1.29, only now with real lenses instead of a diagrammatic representation. (a) Nonimaging optic. (b) Optic with increased angle α_3 at some position results in a decreased angle α_4 somewhere else if the aperture **AB** is fixed.

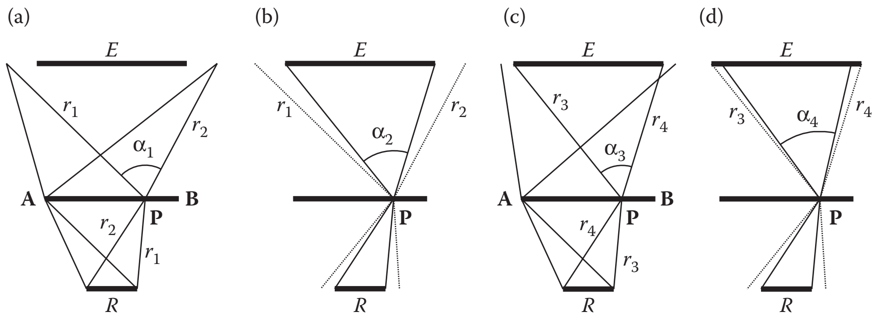


FIGURE 1.31
(a) Ideal nonimaging optic has a wide acceptance angle α_1 wider than *E*. (b) Errors in manufactured nonimaging optic diffuse the light, reducing the acceptance angle to α_2. (c) Nonideal optic has a small acceptance angle α_3. (d) Errors in manufactured nonideal optic further reduce the acceptance angle to α_4 resulting in light loss.

part of the light contained between rays r_1 and r_2 will now miss the receiver, as shown in Figure 1.31b. The acceptance angle is then reduced to α_2, which is still wide enough to capture the light emitted by *E*.

Figure 1.31c shows a nonideal concentrator with aperture **AB**. Since this optic was not designed with the edge ray principle, it has a smaller acceptance angle α_3 at position **P**. When manufactured, the light contained between rays r_3 and r_4 will be diffused as it crosses the optic, resulting in a reduced acceptance angle α_4, as shown in Figure 1.31d. Now the optic is no longer able to capture all the light from emitter *E* and some rays, including r_3 and r_4, will miss the receiver *R* reducing efficiency.

It may then be concluded that the nonimaging optic in Figure 1.31a is more tolerant to errors and maintains efficiency. Even with errors, all light from the emitter is still sent to the receiver. However, the nonideal optic in Figure 1.31c is not as tolerant to errors because it does not maintain efficiency. With errors, some light will miss the receiver, lowering efficiency.

1.7 Nonuniform Sources

When the emitter *E* is uniformly emitting from all its points, the illumination on a receiver *R* will be quite uniform, as shown in Figure 1.32.

However, emitters are not always uniform. They may, for example, be made of several individual small sources. Or there may be tolerances in assembling the light sources, which makes the position of the source vary. Figure 1.33 shows that possibility in which an emitter *E*, when assembled in a real device, may be placed anywhere inside a position tolerance *t*. Depending

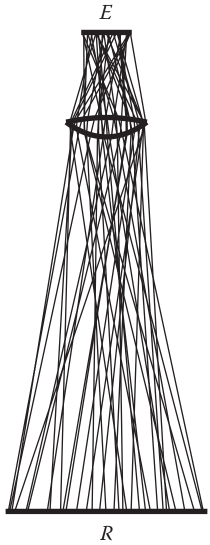


FIGURE 1.32
If the emitter E is fully lit and uniform, the optic creates a quite uniform illumination on receiver R.

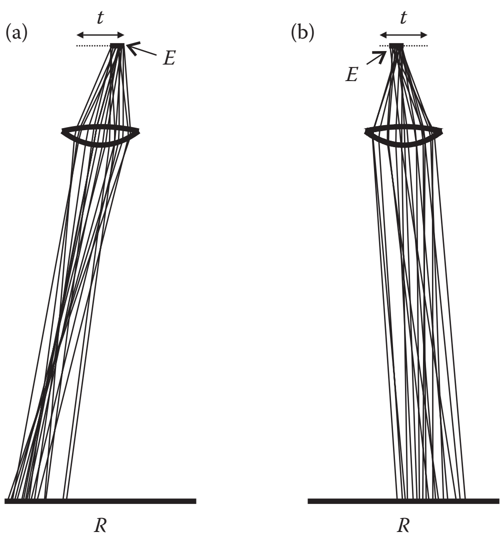


FIGURE 1.33
Due to the manufacturing tolerances, emitter E may be positioned anywhere inside position tolerance t. However, this uncontrolled positioning of emitter E results in an uncontrolled non-uniform illumination of receiver R. (a) and (b) show light crossing the optic for two possible positions of emitter E inside a position tolerance t.

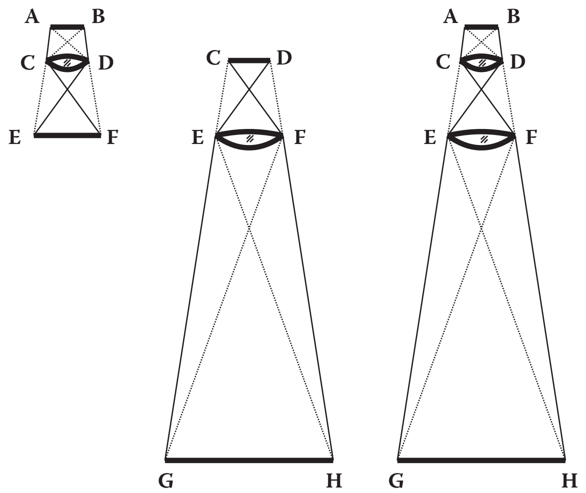


FIGURE 1.34
Left: nonimaging optic **CD** with emitter **AB** and receiver **EF**. Middle: nonimaging optic **EF** with emitter **CD** and receiver **GH**. Right: the two previous optics combined into a device with emitter **AB** and receiver **GH**.

on the position of E inside tolerance t, the illuminance on the receiver may vary considerably. This may be an undesirable characteristic of the collimator optic.

A possible way to avoid this is to combine two nonimaging optics in series, as shown in Figure 1.34.[4] A nonimaging optic **CD** (Figure 1.34 left) has emitter **AB** and receiver **EF**. Another nonimaging optic **EF** (Figure 1.34 middle) has emitter **CD** and receiver **GH**. Combining these two optics (Figure 1.34 right), results in a device with emitter **AB** and receiver **GH**.

The behavior of the resulting device is shown in Figure 1.35. Due to assembly tolerances of a real device, a small emitter E has a placement tolerance t extending from **A** to **B**. Ray r_1 coming from an emitter at position E_1 crosses optic **CD** at point **C**. Nonimaging optic **EF** is designed for a source with edges **CD** and, therefore, it redirects this ray to the edge **H** of receiver R. Accordingly, ray r_2 coming from an emitter at position E_1 crosses optic **CD** at point **D**. Nonimaging optic **EF** redirects this ray to the edge **G** of receiver R. Also, any ray r_3 crossing **EF** coming from inside **CD** will be redirected to a point inside **GH**. These rays have a path similar to ray r_3 in Figure 1.23a, in which a ray coming from a point **P** inside $\mathbf{E}_1\mathbf{E}_2$ is redirected toward a point **Q** inside $\mathbf{R}_1\mathbf{R}_2$. Since an emitter at position E_1 fully illuminates **CD**, there will be rays crossing all points between **C** and **D**. These rays will be spread out by optic **EF** inside **GH**, fully illuminating it.

Even if the emitter moves to other positions, E_2 or E_3, inside position tolerance t, the illumination pattern on R will be approximately maintained, as shown in Figure 1.35.

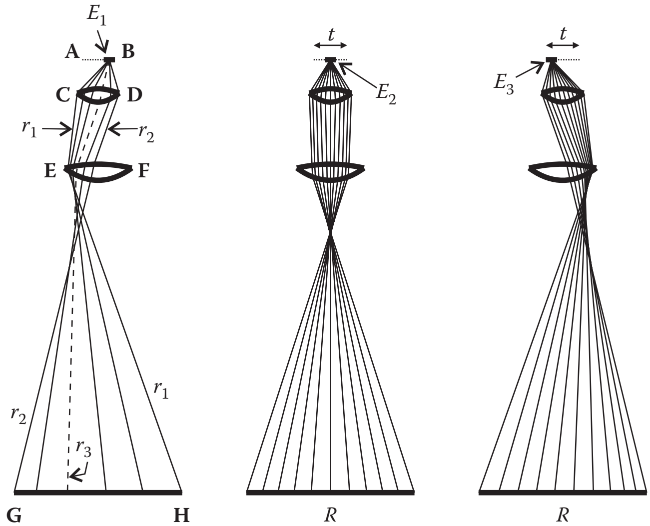


FIGURE 1.35
The emitter may be at different positions E_1, E_2, or E_3 inside position tolerance t and still the optic will fully illuminate the receiver R, extending from **G** to **H**.

Another option for this configuration is color mixing. Figure 1.36 shows a combination of an optic **CD** designed for an emitter **AB** and receiver **EF**, and another optic **EF** with emitter **CD** and receiver **GH**. This system is now used with an emitter composed of three different light sources. These may be, for example, a red source E_R, a green source E_G, and a blue source E_B, forming an RGB (Red, Green, Blue) emitter. When only one of these sources is lit, the receiver R is illuminated by the light of that color. However, when the three are lit simultaneously, these three colors superimpose on the receiver, which will be illuminated by white light. Also, sources E_R, E_G, and E_B may be moved inside position tolerance t (stretching from **A** to **B**), and the optic will still maintain the illumination on receiver R quite constant.

The concepts presented above can also be applied to concentrators. Figure 1.40 shows that for sunlight incidence, angles inside the acceptance angle of the concentrator light is captured and reaches the receiver. However, the irradiance on the receiver may be very nonuniform, with all the light concentrated onto a small spot inside the receiver. This may be an undesirable characteristic of the concentrator optic.

A possible way to avoid this is to combine two nonimaging optics in series, as shown in Figure 1.37. A nonimaging optic **CD** (Figure 1.37 left) has acceptance angle 2θ and receiver **EF**. Another nonimaging optic **EF** (Figure 1.37 middle) has emitter **CD** and receiver **GH**. Combining these two optics (Figure 1.37 right) results in a device with acceptance angle 2θ and receiver **GH**.

The behavior of the resulting device is shown in Figure 1.38. Even for different incidence angles of sunlight inside the acceptance angle, the whole

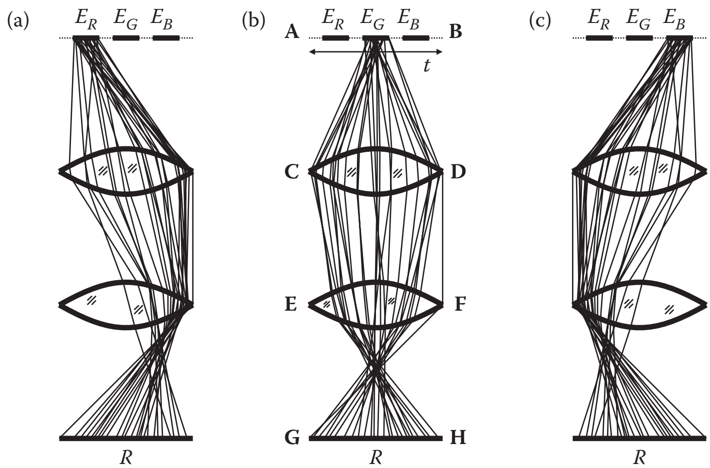


FIGURE 1.36
Suppose that E_R, E_G, and E_B represent light sources of different colors (e.g., red, green, and blue). If only one source is lit, the receiver will be illuminated by the light of that color. If all sources are lit simultaneously, the receiver will be illuminated by white light resulting from mixing all colors. Sources E_R, E_G, and E_B may move inside position tolerance t extending from **A** to **B**. (a), (b), and (c) show, respectively, the situations in which the red, green, or blue light sources are on.

receiver R is always illuminated and the irradiance is quite uniform. The way this optic works is similar to the one in Figure 1.35.

An incoming ray r_1 inside the acceptance angle of optic **CD** will be redirected toward a point inside its receiver **EF**, as shown in Figure 1.38b. Now, if ray r_1 crosses edge **C** of **CD** (the emitter for optic **EF**), it will be redirected by optic **EF**

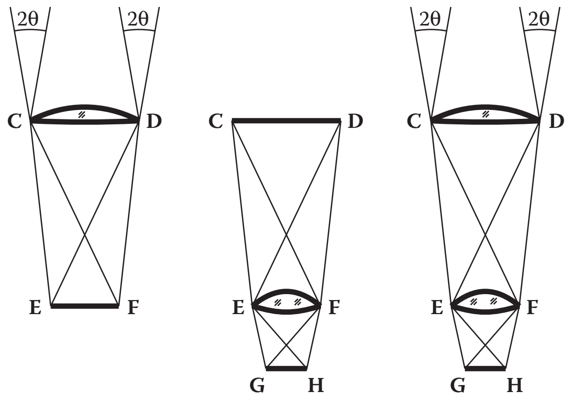


FIGURE 1.37
Left: nonimaging optic **CD** with acceptance angle 2θ and receiver **EF**. Middle: nonimaging optic **EF** with emitter **CD** and receiver **GH**. Right: the two previous optics combined into a device with acceptance angle 2θ and receiver **GH**.

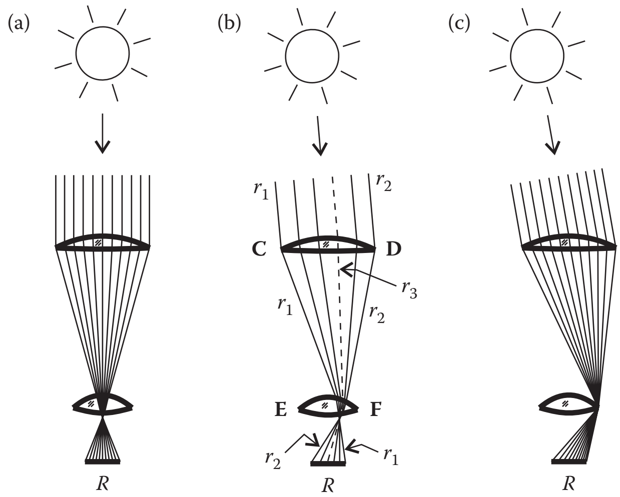


FIGURE 1.38
The receiver R is fully illuminated for different incidence angles of sunlight inside the acceptance angle. (a), (b), and (c) show the paths of light through the system for three different directions of incoming sunlight.

to the edge of receiver R (the receiver of optic **EF**). Something similar happens with another ray r_2. However, since this ray now crosses optic **CD** at point **D**, it is redirected by optic **EF** to the other edge of R. Finally, another ray r_3 inside the acceptance angle of optic **CD** is also redirected by this optic to a point inside its receiver **EF**. Now, for optic **EF**, ray r_3 is coming from a point inside its emitter **CD** and, therefore, it is redirected by optic **EF** to a point inside its receiver R.

1.8 Solar Concentrators

An important application of nonimaging optics is as solar energy concentrators. Figure 1.39a shows a concentrator with entrance aperture a_A and exit aperture (receiver) a_2. The acceptance angle is α_A and the half-acceptance angle is $\theta_A = \alpha_A/2$. This concentrator accepts radiation coming within an angle $\pm\theta_A$ relative to the vertical. Its geometrical concentration is defined as $C_A = a_A/a_2$. Figure 1.39b shows another concentrator with a larger entrance aperture a_B and the same exit aperture (receiver) a_2. Due to the conservation of étendue, since the area of the entrance aperture increased from a_A to a_B, the acceptance angle decreased from α_A to $\alpha_B = 2\theta_B$. This concentrator accepts radiation coming within an angle $\pm\theta_B$ relative to the vertical. Its geometrical concentration is also defined as $C_B = a_B/a_2$, but is now larger than C_A. Therefore, the optic in Figure 1.39a has low concentrations with a wide

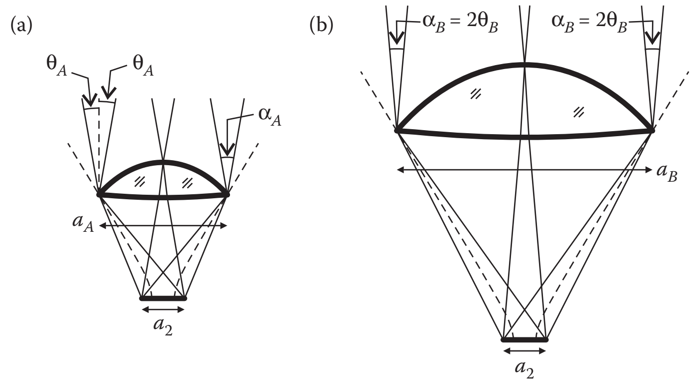


FIGURE 1.39
(a) Low concentration optic with small aperture a_A receiver a_2 and wide acceptance angle $\alpha_A = 2\theta_A$. (b) High concentration optic with large aperture a_B, same receiver a_2 but small acceptance angle $\alpha_B = 2\theta_B$.

acceptance angle, while the optic in Figure 1.39b has high concentration with a small acceptance angle.

We now look at the behavior of an optic with a wide half-acceptance angle θ (low concentration) when exposed to sunlight. In this example, the concentrator is static. As the sun moves across the sky, it first makes an angle $|\beta| > \theta$ to the axis of the concentrator, as shown in Figure 1.40 left. Sunlight enters the optic but fails the receiver, and no light is captured. Eventually, sunlight will make an angle θ to the axis of the concentrator, and light finally reaches the receiver, being captured. As time goes on, sunlight makes an angle $|\beta| < \theta$ to the axis of the concentrator (Figure 1.40 center), and light continues to reach the receiver, being captured. Eventually, sunlight will again make an angle θ to the axis of the concentrator, and then again a larger angle $\beta > \theta$, in which case sunlight again fails to reach the receiver (Figure 1.40 right).

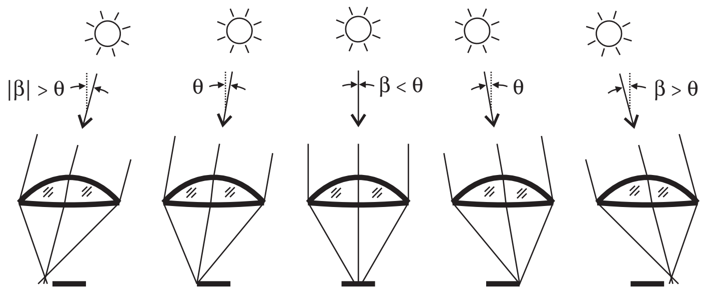


FIGURE 1.40
For a static solar concentrator, as the sun moves across the sky, sunlight is captured if its incidence angle β to the axis of the optic is less than the half-acceptance angle θ of the concentrator.

Nonimaging optics maximize the half-acceptance angle θ and, therefore, also maximize the time a solar concentrator can remain static, while capturing sunlight.

We now look at the behavior of an optic with a small half-acceptance angle θ (high concentration) when exposed to sunlight. Since now the acceptance angle is small, the concentrator can only remain static for a short period of time while capturing sunlight. For that reason, the concentrator must track the sun, and must be constantly orientated to point at the sun, as shown in Figure 1.41. In real systems, however, there will be errors when pointing at the sun, due to imprecision of the tracking system. For a given direction *s* of sunlight, if the axis *b* of the concentrator makes an angle to *s* contained between ±θ, sunlight will still be captured. Therefore, the concentrator has a tolerance of ±θ when pointing at the sun.

Since nonimaging optics maximize the half-acceptance angle θ, they also maximize the tolerance of a concentrator to deviations (errors) when pointing at the sun.

In some applications, several solar concentrators are assembled in modules, as shown in Figure 1.42. In this example, the module is made of three concentrators. The module points in direction **u** but, due to assembly errors, individual concentrators point in slightly different directions $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$.

The acceptance angle of each individual concentrator is α but the module assembly has a smaller acceptance angle α_M. It results from the intersections of the acceptance angles of the individual concentrators. Only within a narrow angle α_A, all the three concentrators are able to simultaneously capture the radiation coming from direction **u**.

As illustrated in Figure 1.40, an ideal concentrator will capture all light if the incidence angle β is inside its half-acceptance angle θ, that is $|\beta| < \theta$. Also, light will miss the receiver if the incidence angle β is outside the acceptance angle $|\beta| > \theta$. The transmission curve is then as shown in Figure 1.43. It is

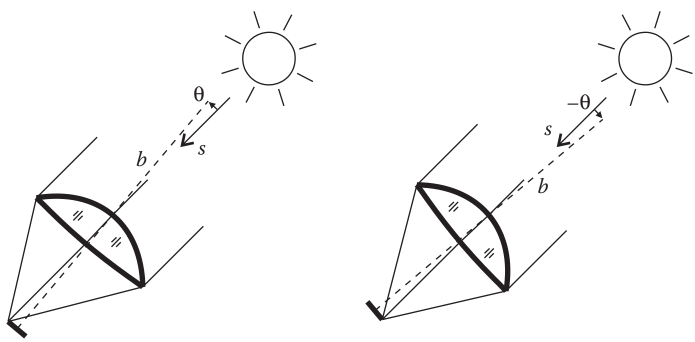


FIGURE 1.41
When pointing at the sun, a solar concentrator has a tolerance in the pointing direction of ±θ where θ is its half-acceptance angle.

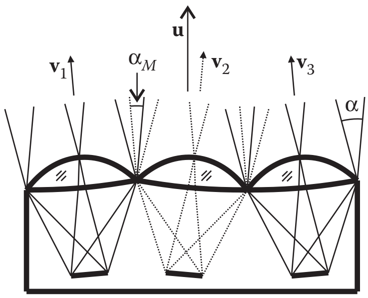


FIGURE 1.42
Module made of three individual concentrators, each with acceptance angle α. Due to assembly errors, individual concentrators point in different directions $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, reducing the acceptance angle of the module to α_M.

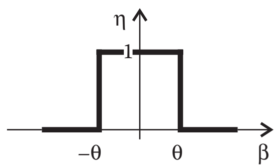


FIGURE 1.43
An ideal concentrator accepts light for incidence angles $|\beta| < \theta$ and rejects light for incidence angles $|\beta| > \theta$ resulting in a stepped transmission (efficiency) curve $\eta(\beta)$.

zero (no light reaches the receiver) for $|\beta| > \theta$, and unity (all light reaches the receiver) for $|\beta| < \theta$.

This result, however, assumes that the incident light is made of parallel rays *r*, as shown in Figure 1.44a. Sunlight, however, has a small angular aperture α_S as shown in Figure 1.44b.

When the finite angular aperture of real sunlight is taken into consideration, the transmission curve of an ideal concentrator changes. Figure 1.45 shows the behavior of a concentrator under real sunlight, with angular aperture α_S.

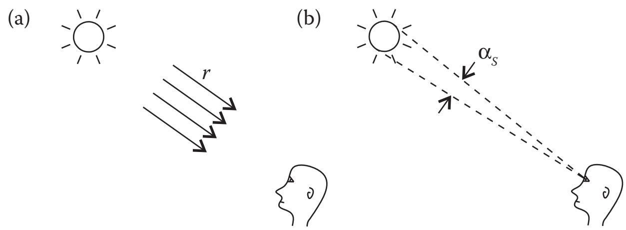


FIGURE 1.44
(a) In some situations, sunlight may be approximated as parallel rays. (b) Real sunlight has a small angular aperture α_S.

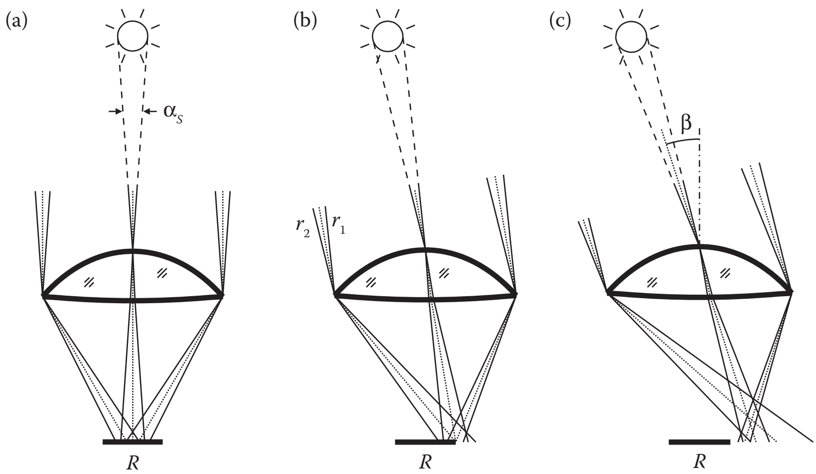


FIGURE 1.45
Sunlight has a finite (although small) angular aperture α_S. (a) All sunlight is captured. (b) Some sunlight is captured and some is not resulting in a smooth transition from capturing to not capturing sunlight. (c) Sunlight is not captured.

When sunlight reaches the concentrator along its axis (as in Figure 1.45a), all light is captured by receiver *R*. As the incidence angle increases, a situation is reached (Figure 1.45b) in which some light reaches the inside of *R* (rays parallel to r_1), and some light misses the receiver (rays parallel to r_2). As the incidence angle β increases still further, all light will miss the receiver, as in Figure 1.45c. This results in a transmission curve with a smooth transition from all light being captured to all light being lost. Figure 1.46 shows the transmission (efficiency) curve η(β) for different incidence angles β of

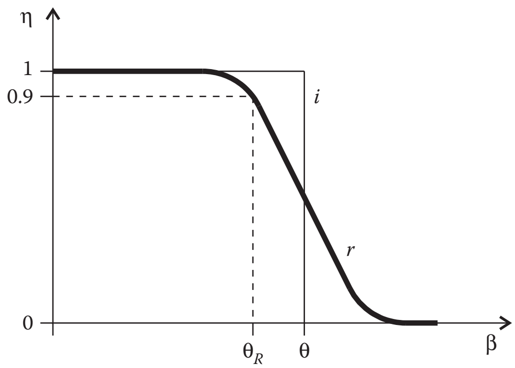


FIGURE 1.46
Ideal transmission curve *i* for different directions of incoming parallel rays and real transmission curve *r* for different directions of incoming sunlight which has a finite angular aperture.

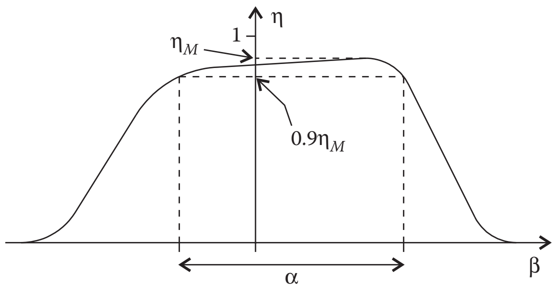


FIGURE 1.47
Transmission curve of an asymmetrical optic with internal losses. The maximum efficiency is η_M and the acceptance angle α is defined at an efficiency of 0.9 η_M.

parallel radiation (ideal curve *i*), and for different incidence angles β of sunlight, which has a finite angular aperture (real curve *r*).

Under parallel radiation, the concentrator has a half-acceptance angle θ. Under sunlight with finite angular aperture, the concentrator has a (reduced) real half-acceptance angle θ_R commonly defined as the angle for which the transmission (efficiency) of the concentrator drops to 90% of its maximum.[3,5]

A more general situation occurs when the optic is asymmetric, and has internal losses. The transmission curve $\eta(\beta)$ is, in this case, also asymmetric, and the maximum efficiency is now a value $\eta_M < 1$, as shown in Figure 1.47. The acceptance angle α is defined at an efficiency of $0.9\eta_M$. In general, the transmission will be a 3-D function of the incidence direction. In that case, the acceptance angle is defined by the circular cone whose acceptance is $0.9\eta_M$.[5] A concentrator is characterized by its maximum efficiency η_M, and its acceptance angle α.

Different imperfections and errors in a concentrator reduce the acceptance angle, as illustrated in Figure 1.48. It shows a concentrator O_1 designed for an acceptance angle α_1. Errors in making the optic (as in Figure 1.24b) reduce its acceptance angle to α_2; errors in assembly (as in Figure 1.42) further reduce the acceptance angle to α_3; other errors and imperfections (tracking errors, wind, dust, overtime wear, and others) further reduce the acceptance angle to a final value α_4. This final acceptance angle value α_4 must still be wider than the sunlight angular aperture α_S, so that all the light coming from the sun is captured.

Starting with a wide acceptance angle α_1 at the design stage is then very important to ensure that the final acceptance angle α_4, of a real device working in the field, is wide enough for the system to have high efficiency.

The acceptance angle may then be seen as a tolerance budget[3] that is spent in different kinds of errors: in the optical surfaces, assembly, installation, tracker structure, sun-tracking, or the finite angular aperture of sunlight.

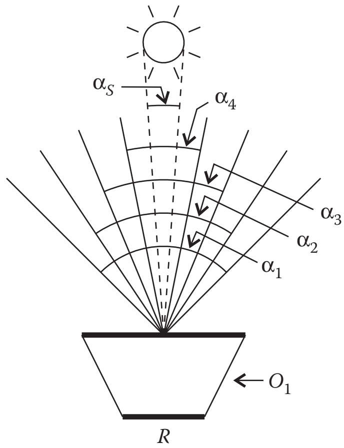


FIGURE 1.48
Solar concentrator O_1 is designed with acceptance angle α_1. Errors in making the optic reduce the acceptance angle to α_2, errors in assembly reduce it to α_3 and other errors and imperfections reduce it to α_4 which must still be wide enough to capture sunlight with angular aperture α_S.

1.9 Light Flux

The amount of light flowing through an aperture depends on how bright the light source is, but also on how much area and angle is available for light to flow through. Figure 1.49 shows light incident on an aperture dx (solid lines). If now we place another aperture dx next to it, doubling the total aperture, the amount of light going through will also double (dotted lines). This means that the amount of light crossing an aperture is proportional to its area: $d\Phi \propto dx$.

Something similar happens relative to angle. Figure 1.50 shows light incident on dx inside an angle $d\theta$ (solid lines). If now we place another cone of light with angle $d\theta$ next to it, doubling the total angle, the amount of light going through will also double (dotted lines). This means that the amount of light crossing an aperture is also proportional to its angle: $d\Phi \propto dx\, d\theta$.

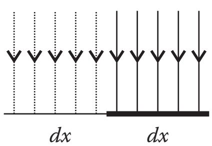


FIGURE 1.49
Doubling the aperture dx, for light to flow through, doubles the amount of light.

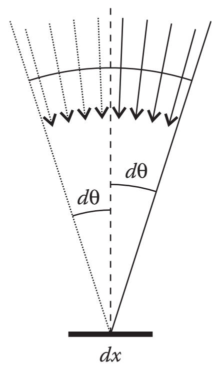


FIGURE 1.50
Doubling the angle $d\theta$ for light to flow through, doubles the amount of light.

Now, if light comes at dx tilted by an angle θ relative to the normal to dx, the projected area of dx in the direction of the incident light is $dx \cos\theta$, as shown in Figure 1.51. The amount of light going through dx is then reduced by a cos θ factor, and is, therefore, $d\Phi \propto dx \cos\theta \, d\theta$.

Finally, light can be brighter or dimmer and the light flux is given by

$$d\Phi = L dx \cos\theta \, d\theta \tag{1.1}$$

where L is the brightness (luminance) of the light.

In three-dimensional geometry, the situation is similar, but now the light flux, instead of being proportional to the plane angle, is proportional to the solid angle. Figure 1.52a shows light incident on dA inside a solid angle $d\Omega = dA_S/r^2$ defined by area dA_S. If now we place another area dA_S next to it, doubling the total solid angle, the amount of light going through will also

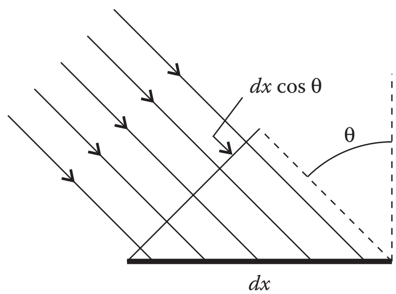


FIGURE 1.51
If the direction of light makes an angle θ to the aperture normal, the projected area in the direction of the incident light is dx cos θ and, therefore, the amount of light crossing dx is proportional to dx cos θ.

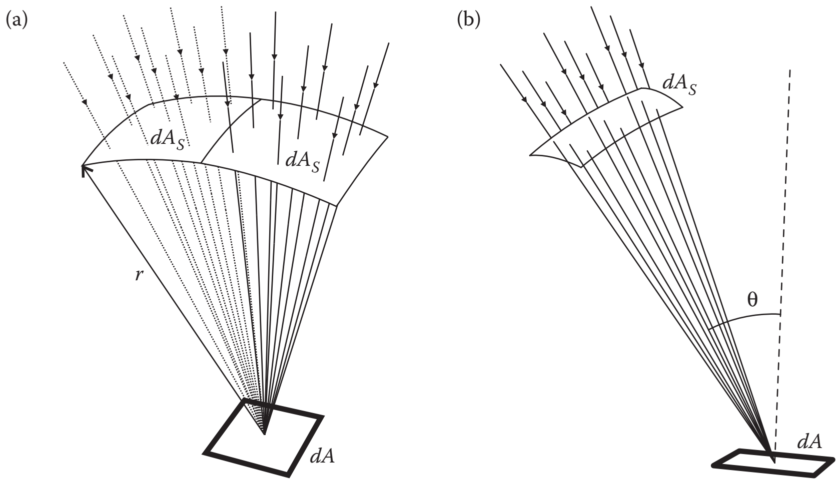


FIGURE 1.52
The solid angle of the light is defined by $d\Omega = dA_S/r^2$. (a) Doubling the area dA_S for light to flow through doubles the solid angle and also doubles the amount of light. (b) The direction of light makes an angle θ to the normal to dA decreasing the amount of light through dA by a factor cos θ.

double (dotted lines). This means that the amount of light crossing an aperture, dA, is proportional to its solid angle in three-dimensional geometry:

$$d\Phi = L dA \cos\theta \, d\Omega$$

where θ is the angle between the direction of light, and the normal to the surface dA, as shown in Figure 1.52b.

Going back to two-dimensional geometry, the flux of light in expression (1.1) may be written as:

$$d\Phi = L dU \tag{1.2}$$

where

$$dU = dx \cos\theta \, d\theta \tag{1.3}$$

is called the étendue of the light (in air or vacuum, where the refractive index is $n = 1$).

If area dx is on the x_1 axis, as in Figure 1.53, this expression becomes:

$$dU = dx_1 \cos\theta_2 \, d\theta_2 \tag{1.4}$$

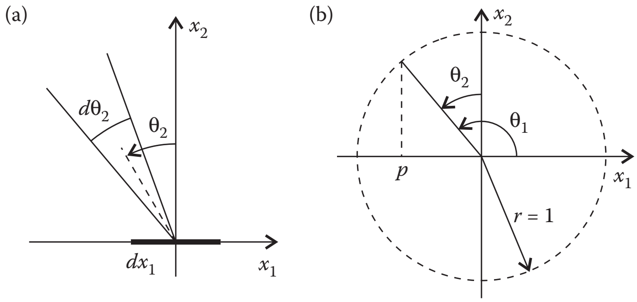


FIGURE 1.53
(a) Light contained within angle $d\theta_2$ crossing area dx_1 on the horizontal axis at an angle θ_2 to its normal (axis x_2). (b) On a circle of unit radius $r = 1$, one has $p = -\sin\theta_2 = \cos\theta_1$.

We may rewrite this expression as

$$dU = dx_1(\cos\theta_2\, d\theta_2) = dx_1\, d(\sin\theta_2) = -dx_1\, dp$$

where $p = -\sin\theta_2 = \cos\theta_1$, as shown in Figure 1.53b (note that in the case in this figure $p < 0$ and $\sin\theta_2 > 0$).

The configuration shown in Figure 1.54 is similar to that in Figure 1.53a, with light crossing dx_1 on the horizontal axis x_1, confined within angle $d\theta_2$ bound by directions $\mathbf{v}_A$ and $\mathbf{v}_B$.

The étendue is in this case given by $dU = -dx_1\, dp = -dx_1\,(p_B - p_A) = dx_1\,(p_A - p_B)$. This expression may now be applied to the étendue of the light exchanged between an emitter **CD** and a receiver **BA** at a height h, as shown in Figure 1.55.

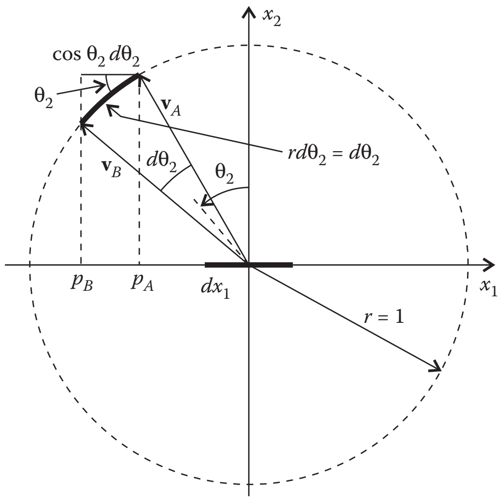


FIGURE 1.54
Light crossing dx_1 is confined between directions $\mathbf{v}_A$ and $\mathbf{v}_B$.

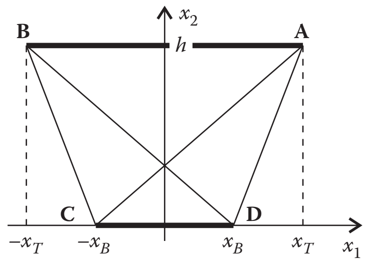


FIGURE 1.55
Emitter **CD** extending from $-x_B$ to x_B and receiver **BA** extending from $-x_T$ to x_T at a height h.

At a point $(x, 0)$ along emitter **CD**, the light reaching receiver **BA** is confined to the cone defined by unit vectors $\mathbf{v}_A$ and $\mathbf{v}_B$ given by

$$\begin{aligned}\mathbf{v}_A &= \frac{\mathbf{A}-(x,0)}{\|\mathbf{A}-(x,0)\|} = \left(\frac{x_T - x}{\sqrt{(x_T-x)^2+h^2}}, \frac{h}{\sqrt{(x_T-x)^2+h^2}}\right)\\ \mathbf{v}_B &= \frac{\mathbf{B}-(x,0)}{\|\mathbf{B}-(x,0)\|} = \left(\frac{-x_T - x}{\sqrt{(-x_T-x)^2+h^2}}, \frac{h}{\sqrt{(-x_T-x)^2+h^2}}\right)\end{aligned} \tag{1.5}$$

The horizontal x_1 coordinates of vectors $\mathbf{v}_A$ and $\mathbf{v}_B$ are given by

$$\begin{aligned}p_A(x) &= \frac{x_T - x}{\sqrt{(x_T-x)^2+h^2}}\\ p_B(x) &= \frac{-x_T - x}{\sqrt{(-x_T-x)^2+h^2}}\end{aligned} \tag{1.6}$$

as shown in Figure 1.56.

The étendue of the light emitted by **CD** and captured by **BA** is then given by

$$U = \int_{-x_B}^{x_B} p_A(x) - p_B(x)dx = \int_{-x_B}^{x_B} p_A(x)dx - \int_{-x_B}^{x_B} p_B(x)dx \tag{1.7}$$

Replacing the values for $p_A(x)$ and $p_B(x)$ from expressions (1.6) we get

$$\begin{aligned}U &= \int_{-x_B}^{x_B} \frac{x_T - x}{\sqrt{(x_T-x)^2+h^2}}dx + \int_{-x_B}^{x_B} \frac{x_T + x}{\sqrt{(x_T+x)^2+h^2}}dx\\ &= \left[-\sqrt{(x_T-x)^2+h^2}\right]_{-x_B}^{x_B} + \left[\sqrt{(x_T+x)^2+h^2}\right]_{-x_B}^{x_B}\\ &= 2\sqrt{(x_T+x_B)^2+h^2} - 2\sqrt{(x_T-x_B)^2+h^2}\end{aligned} \tag{1.8}$$

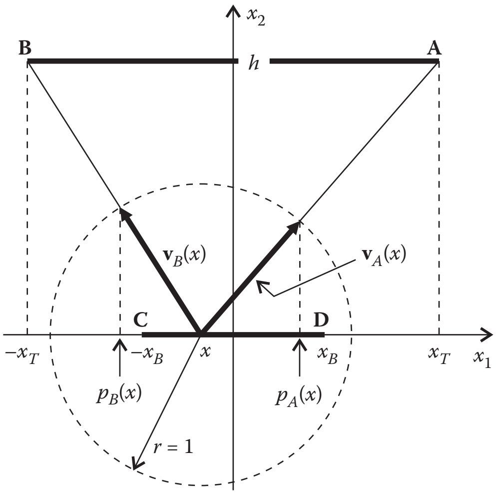


FIGURE 1.56
The étendue (and therefore the light flux) emitted by **CD** that is captured by **BA** is given by integrating the étendue of the light emitted from each point x of **CD** and captured by **BA**.

which can also be written as:

$$U = 2([\mathbf{C},\mathbf{A}] - [\mathbf{D},\mathbf{A}]) \tag{1.9}$$

where [**X**, **Y**] is the distance between points **X** and **Y**.

Referring now to Figure 1.57, points $\mathbf{A}_1$ and $\mathbf{A}_2$ are on hyperbola h with foci **C** and **D**. By being on the hyperbola, points $\mathbf{A}_1$ and $\mathbf{A}_2$ verify $[\mathbf{C}, \mathbf{A}_1] - [\mathbf{D}, \mathbf{A}_1] = [\mathbf{C}, \mathbf{A}_2] - [\mathbf{D}, \mathbf{A}_2]$ and, therefore, the étendue of the light emitted by **CD** and captured by $\mathbf{B}_1\mathbf{A}_1$ is the same as the étendue of the light emitted by **CD** and captured by $\mathbf{B}_2\mathbf{A}_2$.

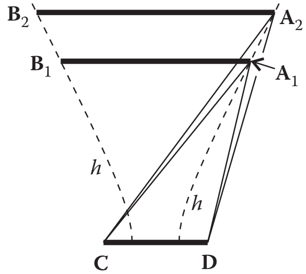


FIGURE 1.57
Points $\mathbf{A}_1$ and $\mathbf{A}_2$ are on a hyperbola h with foci **C** and **D** and, therefore, verify $[\mathbf{C}, \mathbf{A}_1] - [\mathbf{D}, \mathbf{A}_1] = [\mathbf{C}, \mathbf{A}_2] - [\mathbf{D}, \mathbf{A}_1]$. The amount of light exchanged between **CD** and $\mathbf{B}_1\mathbf{A}_1$ is the same as between **CD** and $\mathbf{B}_2\mathbf{A}_2$.

Since the light flux is proportional to the étendue, an optic with aperture $\mathbf{B}_1\mathbf{A}_1$ then captures the same amount of light from emitter **CD** as another optic with aperture $\mathbf{B}_2\mathbf{A}_2$. These hyperbolas are called flow-lines, and are also shown in Figures 1.9 and 1.17 or in Figures 1.21 and 1.22.

1.10 Wavefronts and the SMS

As light travels through an optical system, it may encounter different optical surfaces, where it is refracted or reflected. The optical path length of a ray section between two optical surfaces is defined as the product of the refractive index and the distance the ray travels between those surfaces. The total optical path length of a ray is the sum of the optical path lengths for all ray sections. Figure 1.58 shows a light ray travelling from **P** to **Q** while crossing optical surfaces $\mathbf{c}_1$ and $\mathbf{c}_2$. The refractive index of the material is n_1 between **P** and $\mathbf{c}_1$, n_2 between $\mathbf{c}_1$ and $\mathbf{c}_2$, and n_3 between $\mathbf{c}_2$ and **Q**. Its optical path length from **P** to **Q** is $S = n_1 d_1 + n_2 d_2 + n_3 d_3$.

Figure 1.59 shows a set of light rays, $r_1, r_2, r_3, \ldots$ perpendicular to wavefront w_1. After being deflected at an optical surface $\mathbf{c}(\sigma)$, these rays are now perpendicular to wavefront w_2. If $\mathbf{c}(\sigma)$ is a refractive surface, a ray incident with an angle α_1 to the surface normal emerges at the other side, making an angle α_2 with the surface normal. These angles are related by the law of refraction, which states that $n_1 \sin \alpha_1 = n_2 \sin \alpha_2$, where n_1 is the refractive index before $\mathbf{c}(\sigma)$, and n_2 the refractive index after it.

We now look at the optical path length for two light rays crossing $\mathbf{c}(\sigma)$ at two neighboring points, $\mathbf{C}_1 = \mathbf{c}(\sigma)$ and $\mathbf{C}_2 = \mathbf{c}(\sigma + d\sigma)$, a distance, dc, apart from each other, as shown in Figure 1.60.

The optical path length of the ray through $\mathbf{C}_1$ is $S_1 = n_1 d_7 + n_2(d_{10} + d_8)$ or

$$S_1 = n_1 d_7 + n_2 dc \sin \alpha_2 + n_2 d_8 \tag{1.10}$$

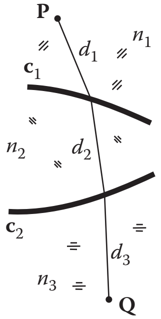


FIGURE 1.58
The optical path length between two points **P** and **Q** is given by $S = n_1 d_1 + n_2 d_2 + n_3 d_3$.

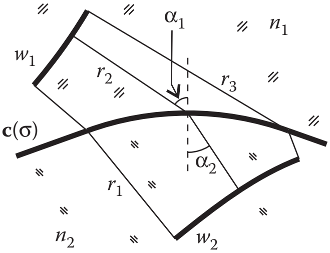


FIGURE 1.59
Light rays r are perpendicular to wavefront w_1, refract at surface $\mathbf{c}(\sigma)$, and emerge at a side which is perpendicular to wavefront w_2.

The optical path length of the ray through $\mathbf{C}_2$ is $S_2 = n_1(d_5 + d_9) + n_2 d_6$ or

$$S_2 = n_1 d_5 + n_1 dc \sin\alpha_1 + n_2 d_6 \tag{1.11}$$

Now, from Figure 1.60, $d_5 = d_7$ and $d_6 = d_8$. Also, from the law of refraction, $n_1 \sin\alpha_1 = n_2 \sin\alpha_2$ and, therefore, $dS = S_2 - S_1 = 0$. Since neighboring rays have the same optical path length between wavefronts w_1 and w_2, all rays between these two wavefronts will also have the same optical path length.

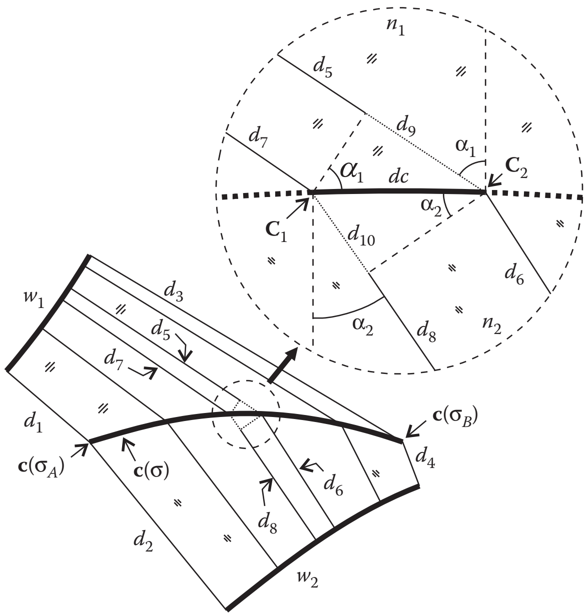


FIGURE 1.60
The optical path length between wavefronts w_1 and w_2 is the same for all light rays between these two wavefronts.

The difference in optical path length for the rays going through $\mathbf{c}(\sigma_A)$ and $\mathbf{c}(\sigma_B)$ is given by

$$S_B - S_A = \int_A^B dS = \int_{\sigma_A}^{\sigma_B} \frac{dS}{d\sigma} d\sigma = 0 \tag{1.12}$$

which means that the optical path length is the same for the rays going through $\mathbf{c}(\sigma_A)$ and $\mathbf{c}(\sigma_B)$. In general, the optical path length is the same for all rays between wavefronts w_1 and w_2.[6,7]

Also, a set of rays perpendicular to a wavefront will remain perpendicular to a wavefront, as the rays travel through an optical system, with refractions and reflections (theorem of Malus and Dupin).[6]

The same conclusion may be obtained for reflection, in which case $n_1 = n_2$. Figure 1.61 shows two neighboring rays reflected on a small portion dc of a mirror. These rays are perpendicular to incoming wavefront w_1 and outgoing wavefront w_2. Here $d_5 + d_{6M} = d_7 + d_{8M}$ and, since $d_8 = d_{8M}$ and $d_6 = d_{6M}$, we get $d_5 + d_6 = d_7 + d_8$ and the optical path length is the same for the two rays reflected at the edges of dc. Note that the distances d_{6M} and d_{8M} and wavefront w_{2M} do not exist physically, they are just a geometrical construction. Just as in the case of refraction, this result may be extrapolated and the optical path length between the wavefronts w_1 and w_2 is the same for all rays between them.

As light travels through multiple optical surfaces, the optical path length is also the same for all rays. Figure 1.62 shows two surfaces, $\mathbf{c}_1$ and $\mathbf{c}_2$, separating three media of refractive indices n_1, n_2, and n_3. Optical path length is the same for all rays between the wavefronts w_1 and w_2, and it is also the same for all rays between wavefronts w_2 and w_3. It must, therefore, also be the same for all rays between the wavefronts w_1 and w_3.

Now, a ray r may be launched from a point $\mathbf{W}_1$ in wavefront w_1 in a direction $\mathbf{v}$, as shown in Figure 1.63. Given the optical path length between the wavefronts w_1 and w_2; this defines the position of point $\mathbf{P}$ along direction $\mathbf{v}$ (perpendicular to w_1), for which, refraction occurs on surface $\mathbf{c}(\sigma)$ separating

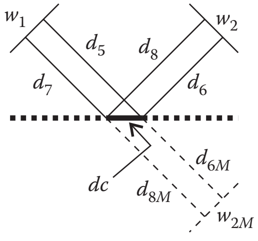


FIGURE 1.61
The optical path length between wavefronts w_1 and w_2 is the same for two neighboring rays reflected on mirror with length dc.

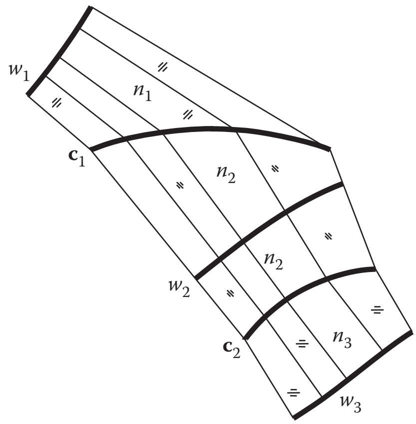


FIGURE 1.62
Optical path length is the same for all rays between the wavefronts w_1 and w_2. It is also the same for all rays between w_2 and w_3 and, therefore, is the same for all rays between w_1 and w_3.

materials of refractive indices n_1 and n_2. That is, there is only one point **P** along **v** for which $S = n_1[\mathbf{W}_1, \mathbf{P}] + n_2[\mathbf{P}, \mathbf{W}_2]$, where S is the optical path length between w_1 and w_2, and $[\mathbf{X}, \mathbf{Y}]$ the distance between points **X** and **Y**. Since we have the direction of the incident and refracted rays at point **P**, we can also calculate the normal $\mathbf{n}_P$ to the surface $\mathbf{c}(\sigma)$ at point **P** (see Chapter 16). Moving point $\mathbf{W}_1$ along wavefront w_1 (rays r_1, r_2, r_3, … in Figure 1.59), with a constant optical path length between w_1 and w_2, allows us to calculate the complete shape of surface $\mathbf{c}(\sigma)$, which is called a Cartesian Oval. The same holds true for reflection, in which case $n_1 = n_2$.

These results may be used for the design of nonimaging optics. Figure 1.64 shows an emitter $\mathbf{E}_1\mathbf{E}_2$ and receiver $\mathbf{R}_1\mathbf{R}_2$, and in between an optic (lens) made of optical surfaces $\mathbf{c}_1$ and $\mathbf{c}_2$. Wavefronts w_1, w_2, w_3, and w_4 are circles centered at $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{R}_1$, and $\mathbf{R}_2$, respectively.

Now, we choose a value for the optical path length S_{14} between wavefronts w_1 and w_4. We also choose an initial point $\mathbf{P}_0$ and its normal $\mathbf{n}_0$ on the top

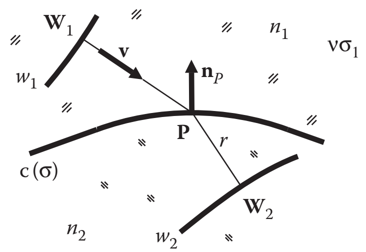


FIGURE 1.63
Given the optical path length between wavefronts w_1 and w_2, this defines the position of point **P** where refraction occurs. The same is valid in the case of reflection (where $n_1 = n_2$).

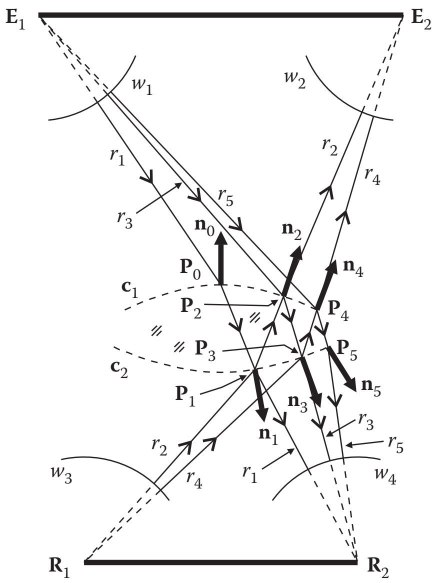


FIGURE 1.64
Constant optical path length between wavefronts w_1 and w_4 and also between w_3 and w_2 allows us to calculate the set of points on a lens focusing $\mathbf{E}_1$ onto $\mathbf{R}_2$ and $\mathbf{E}_2$ onto $\mathbf{R}_1$.

surface $\mathbf{c}_1$ of the lens. Now take the ray r_1 perpendicular to w_1 through $\mathbf{P}_0$. Since we know the normal $\mathbf{n}_0$ at $\mathbf{P}_0$ we may refract r_1 into the lens, calculating the direction of the ray inside the lens (see Chapter 16). Also, since we know the distance ray r_1 travelled between w_1 and $\mathbf{P}_0$, we may calculate the optical path length between $\mathbf{P}_0$ and w_4. With it, it is possible to determine the position of point $\mathbf{P}_1$ and its normal $\mathbf{n}_1$ on the bottom surface $\mathbf{c}_2$ of the lens. Now, from symmetry of the system, the optical path length between w_3 and w_2 is also S_{14}, as between w_1 and w_4. We may now repeat the same process as before. Take the ray r_2, perpendicular to w_3 through $\mathbf{P}_1$. Since we know the normal $\mathbf{n}_1$ at $\mathbf{P}_1$, we may refract r_2 into the lens, calculating the direction of the ray inside the lens. Also, since we know the distance ray r_2 travelled between w_3 and $\mathbf{P}_1$, we may calculate the optical path length between $\mathbf{P}_1$ and w_2. With it, it is possible to determine the position of point $\mathbf{P}_2$ and its normal $\mathbf{n}_2$ on the top surface $\mathbf{c}_1$ of the lens. The process is now repeated with ray r_3 using the same S_{14} optical path length between w_1 and w_4 and a new point $\mathbf{P}_3$ and normal $\mathbf{n}_3$ is obtained on the bottom surface $\mathbf{c}_2$ of the lens. The process is again repeated with ray r_4 using the same S_{14} optical path length between w_3 and w_2, and a new point $\mathbf{P}_4$ and normal $\mathbf{n}_4$ is obtained on the top surface $\mathbf{c}_1$ of the lens. Another ray r_5 allows us to calculate a new point $\mathbf{P}_5$ and corresponding normal $\mathbf{n}_5$ on the bottom surface $\mathbf{c}_2$ of the lens. This process is further repeated, calculating a set of points on both the top and bottom surfaces of the lens simultaneously. This lens will focus point $\mathbf{E}_1$ onto $\mathbf{R}_2$, and $\mathbf{E}_2$ onto $\mathbf{R}_1$.

This is the method used to calculate the lenses in previous figures, such as Figures 1.9b and 1.17b.

References

1. Muñoz, F. et al., Simultaneous multiple surface design of compact air-gap collimators for light-emitting diodes, *Opt. Eng.*, 43, 1522, 2004.
2. Canavarro, C. et al., New second-stage concentrators (XX SMS) for parabolic primaries; Comparison with conventional parabolic trough concentrators, *Sol. Energy*, 92, 98–105, 2013.
3. Pablo, B. et al., High performance Fresnel-based photovoltaic concentrator, *Opt. Express*, 18(S1), 2010.
4. Hernandez, M. et al., High-performance Köhler concentrators with uniform irradiance on solar cell, *Proc. SPIE Vol. 7059, Nonimaging Optics and Efficient Illumination Systems V*, San Diego, California, USA, September 2, 2008.
5. Koshel, R. J., *Illumination Engineering: Design with Nonimaging Optics*, Wiley-IEEE Press, 2013.
6. Born, M. and Wolf, E., *Principles of Optics*, Pergamon Press, Oxford, 1980.
7. Welford, W.T., *Useful Optics*, The University of Chicago Press, Chicago, USA, 1991.

2

Fundamental Concepts

2.1 Introduction

Imaging optical systems have three main components—the object, the optic, and the image it forms. The object is considered as a set of points that emit light in all directions. The light (or part of it) from each point on the object is captured by the optical system and concentrated onto a point in the image. The distances between points on the image may be scaled relative to those on the object, resulting in magnification.

Nonimaging optical systems, instead of an object, have a light source, and instead of an image, have a receiver. Instead of an image of the source, the optic produces a prescribed illuminance (or irradiance) pattern on the receiver.

The first application of nonimaging optics was in the design of concentrators that could perform at the maximum theoretical (thermodynamic) limit. The compound parabolic concentrator (CPC) was the first two-dimensional (2-D) concentrator ever designed, and the success of the device gave birth to nonimaging optics.

This chapter introduces some of the differences between imaging and nonimaging optics, presents the CPC as a concentrator, and shows that it is ideal in two dimensions.

2.2 Imaging and Nonimaging Optics

Figure 2.1 shows a schematic representation of an imaging setup. On the left we have an object **EF**, at the center an optic **CD**, and on the right an image **AB**.

Light coming from edge point **F** on the object must be concentrated onto edge point **A** of the image. Accordingly, light coming from point **E** must be concentrated onto point **B**. This condition would still be valid for any point **P** on the object. Light leaving point **P** is concentrated onto a point **Q** in the

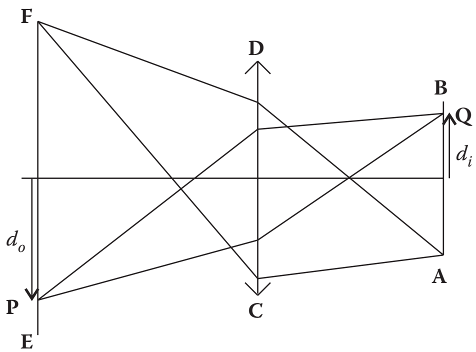


FIGURE 2.1
In an imaging optical system, light coming from any point **P** in the object is concentrated onto a point **Q** in the image, in such a way that $d_i = Md_o$, with d_o and d_i being the distances between **P** and the optical axis, and **Q** and the optical axis, respectively. In particular, light coming from the edge points **E** and **F** of the object, is concentrated onto edge points **B** and **A** of the image, respectively.

image. The distances to the optical axis d_o and d_i from points in the object, and the image, respectively, are related by the following:

$$d_i = Md_o \tag{2.1}$$

where M is the magnification of the system.[1–4] This condition requires that the relative dimensions of several parts of the object are maintained in the image.

Let us now see how to design such a system, using lenses. We can start by concentrating light coming from a point in the object onto the corresponding point in the image. To solve this problem, a Cartesian oval can be used.[1,5] We have, in this case, a set of rays to be focused, and a surface to be defined, as shown in Figure 2.2.

The optical path length along a straight line between **P** and **Q** is given as $S = D + nD_1$. The optical path length of a light ray passing from **P** to **Q** through a point **R** on the surface must also be given by S, so we must have $S = d + nd_1$. This condition enables us to obtain all the points of the Cartesian oval.

If we now want, nevertheless, to focus two points of the object onto two points of the image **AB**, a surface is no longer sufficient. We then need at least two surfaces. Let us then suppose that, in fact, two surfaces are sufficient. We now have two sets of edge rays that are to be focused, those coming from **E** and **F** (that must be focused to **B** and **A** respectively), and we have two surfaces to be defined. Let us then suppose that a lens similar to the one presented in Figure 2.3 can be designed so that it focuses the two sets of edge rays of the object onto the two sets of edge rays of the image (later in this chapter, a way to design such a lens is presented).

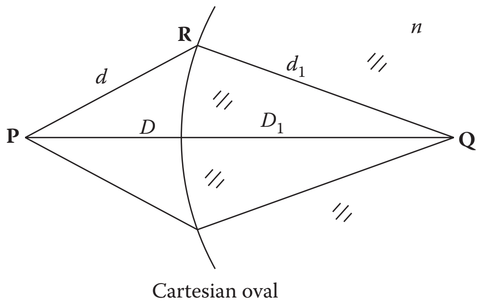


FIGURE 2.2
To solve the problem of forming an image through an optical system, we can start by trying to focus light coming from a point on the image onto a point on the object. A way to achieve this is by using a Cartesian oval. In this case, each point on the surface is crossed by just one ray of light coming from the object. It is then possible to choose the slope of the surface so that convergence is guaranteed.

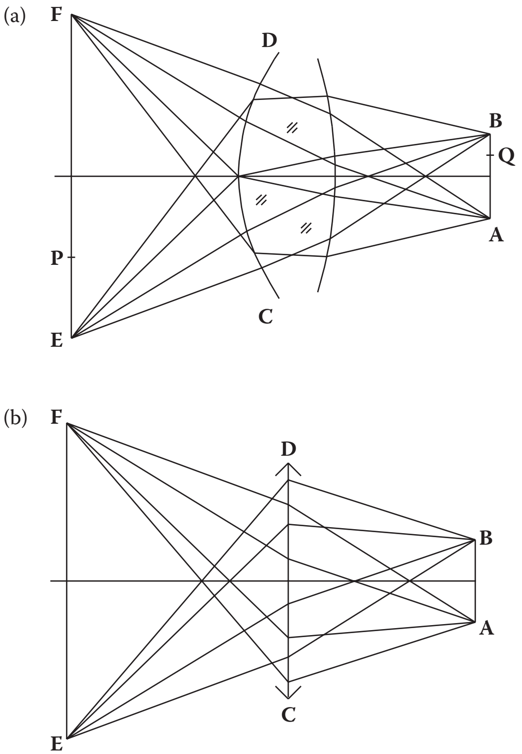


FIGURE 2.3
(a) A lens that focuses onto **A** and **B** the light coming from **F** and **E**, respectively. Note that **E** and **F** are edges of the "object" and that **A** and **B** are edges of the "image," and (b) the same optical system, but in a schematic way.

However, this new lens does not guarantee that light coming from an intermediate point **P** in the object is concentrated onto the corresponding point **Q** in the image, because there are not enough degrees of freedom to do so. To add new degrees of freedom, however, more surfaces must be added. Since a lens can have only two surfaces, more lenses must be added. To guarantee that the light coming from more points in the object is concentrated onto the corresponding points in the image, the systems become more complex. Eventually, this would lead us to systems having an infinite number of lenses.[6,7]

If we do not intend to increase the number of lenses, a new degree of freedom must be found, that allows the focusing of several points of the object onto the corresponding points in the image. One way is to consider a lens whose refractive index varies from point-to-point in its interior.[3,6,7] This kind of solution is, nonetheless, hard to implement, because it is difficult to build a material with a refractive index varying in accordance with the results of the calculations.

Owing to these and other difficulties in designing an ideal imaging device, the optical devices available do not produce perfect images, but images with aberrations. These arguments do not prove that it is impossible to make (build) a perfect imaging system, they only show that this task does not seem to be easy.

Although the lens of Figure 2.3 does not guarantee the formation of an image, it does guarantee that all the radiation exiting **EF** will eventually pass across **AB**. In fact, if the light rays exiting the edges of the source **E** and **F** pass through edges **A** and **B** of the receiver, the light rays exiting intermediate points **P** of the source must also exit between points **A** and **B**. Therefore, in this case, all the radiations coming from **EF** and hitting **CD** will end up concentrated at **AB**. This lens then acts as a concentrator with **EF** as source and **AB** as receiver. This is illustrated in Figure 2.4. In this case, ray r_1 coming from

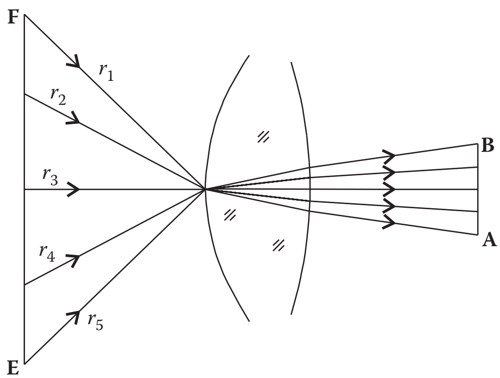


FIGURE 2.4
If ray r_1 coming from the edge **F** of the source is deflected to edge point **A** on the receiver, and ray r_5 coming from the edge **E** of the source is deflected to edge **B** of the receiver, all other rays—r_2, r_3, r_4—coming from intermediate points in source **EF** will end between points **A** and **B** on the receiver.

edge point **F** of the source is deflected toward edge point **A** of the receiver and ray r_5 coming from edge **E** of the source is deflected to edge point **B** of the receiver. Therefore, rays r_2, r_3, and r_4, coming from intermediate points in the source, are deflected to intermediate points on the receiver.

Generally, nevertheless, the light rays coming from a point **P** in the object, as shown in Figure 2.3, will not converge onto a point **Q**, so that no image will be formed at **AB**.

As seen, many degrees of freedom are required for the design of an imaging system, because the formation of an image imposes a large number of conditions that must be fulfilled simultaneously. From these results the difficulty of designing a perfect imaging device, since the number of available degrees of freedom for the design of an optical system is usually not sufficient. If the objective is, nonetheless, just to transfer the energy from a source to a receiver, image formation is unnecessary. Instead, it suffices to require that the light rays coming from the edges of the source are transformed into rays going to the edges of the receiver, as shown in Figure 2.4. Now there are far fewer requirements, and only a small number of degrees of freedom will result in an ideal device.

If the light source is displaced to infinity, becoming infinitely large, the situation presented in Figure 2.3 becomes that of Figure 2.5.

In this case, the incoming radiation can be characterized by the angular aperture θ. This lens now works as a device concentrating onto **AB** all the radiation with half-angular aperture θ falling on **CD**. This device must be designed such that the parallel rays d_1 are concentrated onto **A** and the

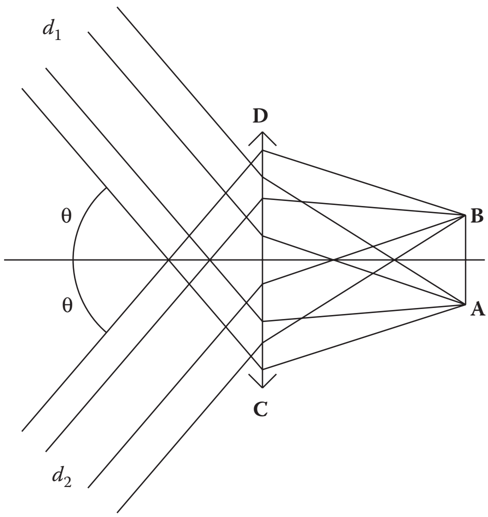


FIGURE 2.5
The limit case of Figure 2.3b, in which the edge points **E** and **F** are displaced to infinity. Now the radiation arriving to the optical system **CD** has an angular aperture θ for each side. Edge rays d_1 are concentrated onto point **A** and edge rays d_2 are concentrated onto point **B**.

parallel rays d_2 are concentrated onto **B**. In this manner, all the radiation falling on the device, making an angle to the optical axis smaller than θ, must pass between **A** and **B**.

We can also compare the optical devices presented in Figure 2.1 and 2.3. In both the cases, the condition is such that the light coming from **EF** must pass through **AB**. In the case of the device presented in Figure 2.1, it is also required that light coming from **F** must be concentrated onto **A**, and that the light coming from **E** must be concentrated onto **B**. Besides, light coming from any other point **P** must be concentrated onto a point **Q** on the image, being the distances d_0 and d_i of **P** and **Q** to the optical axis related by Equation 2.1.

In the case of the device presented in Figure 2.3 the only requirement is that the light coming from **F** must be concentrated onto **A**, and that the light coming from **E** must be concentrated onto **B**. The light coming from a generic point **P** of the object will not be necessarily concentrated onto any point along **AB,** so generally no image will be formed.

The device presented in Figure 2.1 is imaging, and the one presented in Figure 2.3 is nonimaging. Note that both perform the same when used as radiation collectors.

2.3 The Compound Parabolic Concentrator

As described earlier, nonimaging devices can be used as concentrators. In this case, the formation of an image is not a necessary condition. The only condition is that the radiation entering the optical device ends up being concentrated at its exit.

It was mentioned earlier that optical systems have aberrations. As a matter of fact, these can be divided into several categories. The device presented in Figure 2.3 can have, for example, chromatic aberrations.[1,2,8] This nonideality results from the fact that several wavelengths of light are refracted in different directions. One of the best known applications of this effect is the use of prisms to separate white light from the sun, into its several spectral colors. To avoid this aberration, mirrors can be used, because all wavelengths are reflected in the same way.

We start with a radiation source and a receiver onto which we want to concentrate as much light as possible coming out of the source. Figure 2.6a shows a source (emitter) E_1 and a receiver **AB**.

If now this source moves to the left, as shown in Figure 2.6b, and grows in size from E_1 to E_2, so that its edges always touch the rays r_1 and r_2, which make an angle 2θ between each other, the radiation field at **AB** will tend to be the one in Figure 2.7, in which the receiver **AB** is shown in a horizontal orientation. At each point, the receiver **AB** “sees” the incoming radiation contained between two edge rays that make an angle 2θ between each

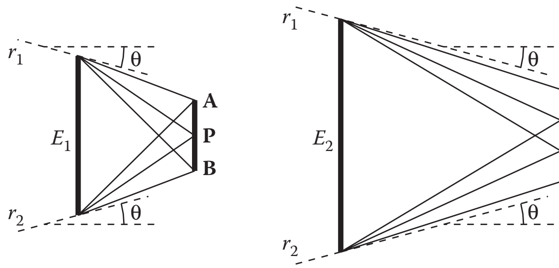


FIGURE 2.6
As the source E moves to the left, and grows so that its edges always touch the rays r_1 and r_2, its size will be E_1, E_2, … The radiation received at **AB** tends to be confined at every point to an angle 2θ.

other. These edge rays are coming from the infinite source E at an infinite distance.

Our goal is to concentrate this radiation to the maximum possible extent, that is, to send the maximum power through the aperture **AB**. Our approach is to let **AB** be the exit aperture of the device, and then generate mirror profiles upward from points **A** and **B**. We may start with simple flat mirrors, placing one on point **A**, and another on point **B**. Owing to the symmetry of the problem about the vertical line through mid point **P**, these mirrors are also symmetrical. This situation is presented in Figure 2.8.

To deflect onto **AB** the maximum possible radiation, angle β must be as small as possible, so that the entrance aperture $\mathbf{C_1D_1}$ can be as large as possible. But there is a limit to the minimum value of β, which is reached when the ray of light r_1, reflected at $\mathbf{D_1}$, is redirected to point **A**. If β is smaller, there will be rays reflected by $\mathbf{BD_1}$ onto $\mathbf{AC_1}$, and from there, away from **AB**. After placing the first mirror, a second one can be added above it. Figure 2.9 presents this possibility.

Also in this case, the slope of the mirrors is chosen so as to maximize the width of the entrance aperture, which is now $\mathbf{C_2D_2}$. Again this means that this mirror must redirect the edge rays coming from the left, so that the ray r_2 is reflected at $\mathbf{D_2}$ toward point **A**. We can now add more and more mirrors

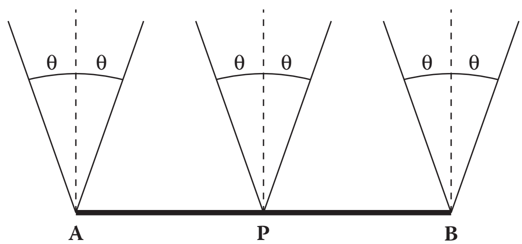


FIGURE 2.7
Uniform radiation of angular aperture θ for each side, and falling on a surface **AB**.

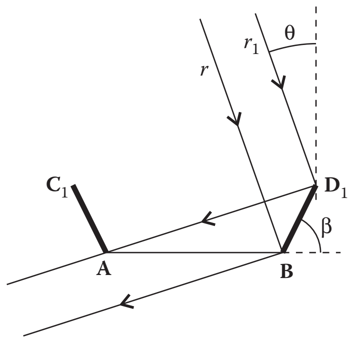


FIGURE 2.8
To concentrate radiation onto **AB**, we can place mirrors at **A** and **B**. To capture the maximum amount of radiation, entrance aperture $\mathbf{C_1D_1}$ must also be a maximum. Therefore, the angle β that these mirrors make with the horizontal must be a minimum. This minimum value of β is obtained when the edge ray r_1 coming from the left, and falling on $\mathbf{D_1}$, is reflected toward **A**. If β decreases, this light ray would be reflected at $\mathbf{D_1}$, then at mirror $\mathbf{AC_1}$, and from there would be reflected away from **AB**. Mirror $\mathbf{AC_1}$ is symmetrical to $\mathbf{BD_1}$.

atop one another. These mirrors have a finite size, but they can be made as small as desired. As this happens, more and more smaller mirrors can be added. The mirrors together tend to adjust to a curve. This situation is presented in Figure 2.10. Angle β, which was minimized previously for each small mirror, is now the slope of the curve and must also be minimized at each point.

Considering the way this curve is defined, it must deflect onto a point **A** the edge rays *r* coming from the left. We then have a curve that deflects a set of parallel rays onto a point. The geometrical curve having this characteristic is a parabola, so that the curve is a parabola with its axis parallel to the edge rays *r* coming from the left, and having its focus at point **A**. It can also be

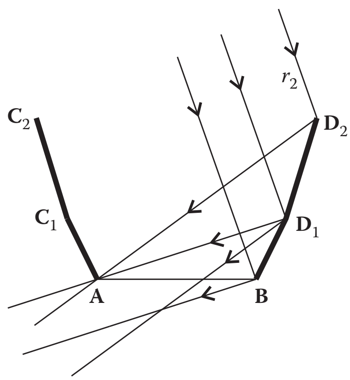


FIGURE 2.9
Using the same method presented in Figure 2.8, it is now possible to add new mirrors at points $\mathbf{C_1}$ and $\mathbf{D_1}$, enlarging even more the dimension of the entrance aperture that now becomes $\mathbf{C_2D_2}$.

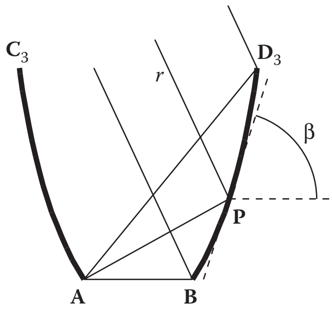


FIGURE 2.10
The procedure presented in Figure 2.9 can now be extended by adding more mirrors and diminishing their size.

noted that this curve is the one that, at each point **P**, produces the smallest value for β, that is, the one that leads to a maximum entrance aperture $\mathbf{C}_3\mathbf{D}_3$.

As can be seen, from Figure 2.11, if the parabola is extended upward, there comes a point where it starts tilting inside, reducing the size of the entrance aperture.

When this happens, the top of the right mirror starts to shadow the bottom of the left, and vice versa. Since we are interested in obtaining the maximum possible entrance aperture, the parabolas must be cut at line **CD** where the distance between them is maximum. The final concentrator must then look like Figure 2.12.

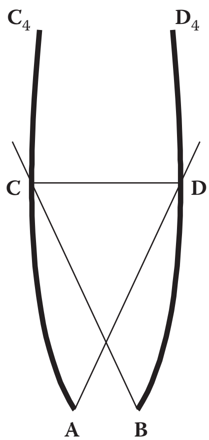


FIGURE 2.11
As the parabolas are extended upward, the distance between the mirrors increases until a maximum **CD** is reached, and then starts to decrease. Also, portions $\mathbf{DD}_4$ and $\mathbf{CC}_4$ of the mirror shadow the other portions of mirror **AC** and **BD**, respectively. Since the goal is to maximize the size of the entrance aperture, the parabolas must be cut at **CD**.