

INTRODUCTION AND TERMINOLOGY

John Koshel

This chapter introduces the reader to a number of terms and concepts prevalent in the field of illumination optics. I establish the units basis that is used throughout this book. The fields of nonimaging and illumination optics have a fundamental basis on these units; therefore, it is demanded that the reader be well versed in units and how to design, analyze, and measure with them. Next, I give an overview of the field and important parameters that describe the performance of an illumination system. The next chapter on étendue expands upon this treatment by introducing terms that are primarily focused on the design of efficient illumination systems.

1.1 WHAT IS ILLUMINATION?

Until recently the field of optical design was synonymous with lens or imaging system design. However, within the past decade, the field of optical design has included the subfield of illumination design. Illumination is concerned with the transfer of light, or radiation in the generic sense,* from the source(s) to the target(s). Light transfer is a necessity in imaging systems, but those systems are constrained by imaging requirements. Illumination systems can ignore the “imaging constraint” in order to transfer effectively the light. Thus, the term nonimaging optics is often used. In the end, one may classify optical system design into four subdesignations:

- *Imaging Systems.* Optical systems with the imaging requirement built into the design. An example is a focal-plane camera.
- *Visual Imaging Systems.* Optical systems developed with the expectation of an overall imaging requirement based upon integration of an observation system. Examples include telescopes, camera viewfinders, and microscopes that require human observers (i.e., the eye) to accomplish imaging.

* Light is defined as that within the visible spectrum, but often it loosely includes the ultraviolet and infrared parts of the electromagnetic spectrum. The proper term is electromagnetic radiation.

- *Visual Illumination Systems.* Optical systems developed to act as a light source for following imaging requirements. Examples include displays, lighting, and extend to the illuminator for photocopiers.
- *Nonvisual Illumination Systems.* Optical systems developed without the imaging criterion imposed on the design. Examples include solar concentrators, optical laser pump cavities, and a number of optical sensor applications.

The latter two systems comprise the field of illumination engineering. Imaging systems can be employed to accomplish the illumination requirements, but these systems are best suited for specific applications. Examples include critical and Köhler illumination used, for example, in the lithography industry, but as this book shows, there are a number of alternative methods based on nonimaging optics principles. This book focuses on these nonimaging techniques in order to transfer light effectively from the source to the target, but imaging principles are used at times to improve upon such principles. Additionally, I place no requirement on an observer within the system, but as you will discover, most illumination optics are designed with observation in mind, including the human eye and optoelectronic imaging, such as with a camera. To neglect the necessary visualization and its characteristics often has a detrimental effect on the performance of the illumination system. This last point also raises the subjective perception of the illumination system design. This factor is not currently a focus of this book, but it is discussed in order to drive the development of some systems.

I use the remainder of this chapter to discuss:

- A short history of the illumination field
- The units and terminology for illumination design and analysis
- The important factors in illumination design
- Standard illumination optics
- The steps to design an illumination system
- A discussion of the difficulty of illumination design, and
- The format used for the chapters presented herein.

Note that I typically use the terms illumination and nonimaging interchangeably, but, in fact, illumination is a generic term that includes nonimaging and imaging methods for the transfer of light to a target.

1.2 A BRIEF HISTORY OF ILLUMINATION OPTICS

The history of the field of illumination and nonimaging optics is long, but until recently it was mostly accomplished by trial and error. Consider Figure 1.1, which shows a timeline of the development of sources and optics for use in the illumination field [1]. Loosely, the field of illumination optics starts with the birth of a prevalent light source on earth—the Sun. While the inclusion of the Sun in this timeline may at first appear facetious, the Sun is becoming of increasing importance in the illu-

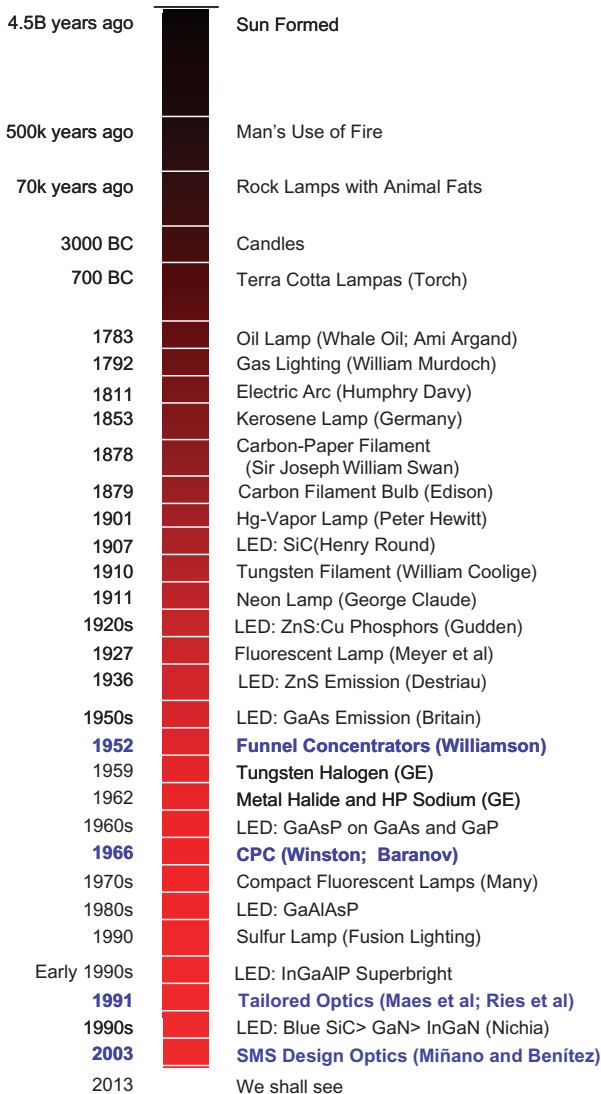


Figure 1.1 A timeline of the history of illumination and nonimaging optics. On the left is the approximate date of inception for the item on the right. Items listed in blue are illumination optic design concepts, those leading with LED are new solid-state lighting sources, and the remainder is other types of sources.

mination and nonimaging optics communities. This importance is borne out of daylighting systems, solar thermal applications, and solar energy generation. The use, modeling, and fabrication of sources is one of the largest components of the field of illumination design. Increasingly, LED sources are supplanting traditional sources since LEDs provide the potential for more efficient operation, color selection, long lifetimes, and compact configurations. It is only in the past 60 years that

nonimaging optical methods have been developed. The illumination industry, both in design and source development, is currently burgeoning, so vastly increased capabilities are expected in the next few decades.

1.3 UNITS

As with any engineering or scientific discipline, the use of units is imperative in the design and modeling of illumination systems. It is especially important to standardize the system of units in order to disseminate results. There are essentially two types of quantities used in the field of illumination:

- *Radiometric Terms.* Deterministic quantities based on the physical nature of light. These terms are typically used in nonvisual systems; and
- *Photometric Terms.* Quantities based on the human visual system such that only visible electromagnetic radiation is considered. This system of units is typically used in visual systems.

Radiometric and photometric quantities are connected through the response of eye, which has been standardized by the International Commission on Illumination (Commission Internationale de L'Éclairage; CIE) [2, 3]. Both of these set of terms can be based on any set of units, including English and metric; however, standardization at the Fourteenth General Conference on Weights and Measures in 1971 on the metric system is defined by the International System (Système Internationale; SI) [4]. The units for length (meter; m), mass (kilogram; kg), and time (second; s) provide an acronym for this system of units: MKS. There is an analogous one that uses the centimeter, gram, and second, denoted as CGS. This book uses the MKS standard for the radiation quantities, though it often makes use of terms, especially length, in non-MKS units, such as the millimeter. In the next two subsections, the two set of terms are discussed in detail. In the section on photometric units, the connection between the two systems is presented.

1.3.1 Radiometric Quantities

Radiometry is a field concerned with the measurement of electromagnetic radiation. The radiometric terms as shown in Table 1.1 are used to express the quantities of measurement* [5]. The term radiant is often used before a term, such as radiant flux, to delineate between like terms from the photometric quantities; however, the accepted norm is that the radiometric quantity is being expressed if the word radiant is omitted. Additionally, radiometric quantities are often expressed with a subscript “e” to denote electromagnetic. Omission of this subscript still denotes a radiometric quantity.

* Note that I follow the convention of R.W. Boyd, *Radiometry and the Detection of Optical Radiation* for the convention of the symbols. Be cautious and forewarned that there are different conventions for the symbols for the radiometric and photometric quantities, which can cause confusion for both beginning and advanced readers.

TABLE 1.1 Radiometric Terms and their Characteristics

Term and description	Symbol	Functional form	SI units
Radiant energy	Q_e		J
Radiant energy density	u_e	$\frac{dQ_e}{dV}$	J/m ³
Radiant energy per unit volume			
Radiant flux/power	Φ_e or P_e	$\frac{dQ_e}{dt}$	J/s or W
Radiant energy per unit time			
Radiant exitance	M_e	$\frac{d\Phi_e}{dA_{\text{source}}}$	W/m ²
Radiant flux per unit source area			
Irradiance	E_e	$\frac{d\Phi_e}{dA_{\text{target}}}$	W/m ²
Radiant flux per unit target area			
Radiant intensity	I_e	$\frac{d\Phi_e}{d\Omega}$	W/sr
Radiant flux per unit solid angle			
Radiance	L_e	$\frac{d^2\Phi_e}{dA_{\text{s,proj}}d\Omega}$	W/m ² /sr
Radiant flux per unit projected area per unit solid angle			

Radiometric quantities are based on the first term in the table, radiant energy (Q_e), which is measured in the SI unit of joules (J). The radiant energy density (u_e) is radiant energy per unit volume measured in the SI units of J/m³. The radiant flux or power (Φ_e or P_e) is the energy per unit time, thus it is measured in the SI unit of J/s or watts (W). There are two expressions for the radiant surface flux density, the radiant exitance and the irradiance. The radiant exitance (M_e) is the amount of flux leaving a surface per unit area, while the irradiance (E_e) is the amount of flux incident on a surface per unit area. Thus, the exitance is used for source emission, scatter from surfaces, and so forth, while irradiance describes the flux incident on detectors and so forth. Both of these terms are measured in SI units of watts per square meter (W/m²). Radiant intensity (I_e) is defined as the power radiated per unit solid angle, thus it is in the SI units of watts per steradian (W/sr). Note that many describe the radiant intensity as power per unit area (i.e., radiant flux density), but this is expressly incorrect.* Most texts use the term of intensity when irradiance is the correct terminology, but some excellent texts use it correctly, for example, *Optics* by Hecht [6]. Modern texts in the field of illumination use it correctly [7]. For a thorough discussion on this confusion, see the article by Palmer [8]. Finally, the radiance (L_e) is the power per unit projected source area per unit solid angle, which is in the SI units of watts per meter squared per steradian (W/m²/sr). In Sections 1.4–1.6, the quantities of irradiance, intensity, and radiance, respectively, are discussed in more detail.

* The use of the term intensity when irradiance is implied is unfortunate and causes considerable confusion, especially to those learning optics or radiometry. Some authors denote that “intensity” is the power per unit area, and that “radiant intensity” is power per unit solid angle. Others indicate that in physical optics intensity is defined as the magnitude of the Poynting vector. These positions are inaccurate due to the SI definition, and in my opinion should be corrected.

When one of the quantities listed in Table 1.1 is provided as a function of wavelength, it is called a spectral quantity. For example, when the intensity has a spectral distribution, it is called the spectral radiant intensity. The notation for the quantity is modified with either a λ subscript (e.g., $I_{e,\lambda}$) or by denoting the quantity is a function of wavelength (e.g., $I_e(\lambda)$). The units of a spectral quantity are in the units listed in Table 1.1, but are per wavelength (e.g., nm or μm). In order to compute the value of the radiometric quantity over a desired wavelength range, the spectral quantity is integrated over all wavelengths

$$f_e = \int_0^{\infty} h(\lambda) f_e(\lambda) d\lambda, \quad (1.1)$$

where $h(\lambda)$ is the filter function that describes the wavelength range of importance, f_e is the radiant quantity (e.g., irradiance), and $f_e(\lambda)$ is the analogous spectral radiant quantity (e.g., spectral irradiance).

1.3.2 Photometric Quantities

The photometric terms are applied to the human visual system, so only the visible spectrum of 360–830 nm adds to the value of a term. Due to the variability of the human eye, a standard observer is used, which is maintained by the CIE. Table 1.2 shows the analogous quantities to that of the radiometric terms of Table 1.1. The term luminous is used before the term, such as luminous flux, to delineate between radiometric and photometric quantities. Additionally, radiometric quantities are expressed with a subscript “v” to denote visual.

Luminous energy (Q_v) is measured in the units of the talbot (T), which is typically labeled as lumen-s (lm-s). The luminous energy density (u_v) is in the units of

TABLE 1.2 Photometric Terms and Their Characteristics

Term and description	Symbol	Functional form	Units
Luminous energy	Q_v		T
Luminous energy density	u_v	$\frac{dQ_v}{dV}$	T/m ³
Luminous flux/power	Φ_v or P_v	$\frac{dQ_v}{dt}$	lm
Luminous exitance	M_v	$\frac{d\Phi_v}{dA_{\text{source}}}$	lx
Illuminance	E_v	$\frac{d\Phi_v}{dA_{\text{target}}}$	lx
Luminous intensity	I_v	$\frac{d\Phi_v}{d\Omega}$	cd
Luminance	L_v	$\frac{d^2\Phi_v}{dA_{s,\text{proj}}d\Omega}$	nt

T/m^3 . Once again, $\text{lm}\cdot\text{s}$ is typically used for the talbot. The luminous flux is provided in the units of lumens (lm). The two luminous flux surface density terms are for a source, the luminous exitance (M_v), and for a target, the illuminance (E_v). Both terms have the units of lux (lx), which is lumens per meter squared (lm/m^2). The luminous intensity (I_v) is measured in candela (cd), which is lumens per unit steradian (lm/sr). Note that the candela is one of the seven SI base units [9]. The definition for the candela was standardized in 1979, which per Reference [9] states that it is “the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of $1/683$ watt per steradian.” This definition may appear arbitrary, but it is established on previous definitions, thus provides a degree of consistency over the lifetime of the candela unit. The standardization of the candela for luminous intensity provides further reason of the need to correct the misuse of the terms intensity and irradiance/illuminance. The luminance (L_v) is the photometric analogy to the radiometric radiance. It is in the units of the nit (nt), which is the lumens per meter squared per steradian ($\text{lm/m}^2/\text{sr}$). There are several other photometric units for luminance and illuminance that have been used historically. Table 1.3 provides a list of these quantities. Note that these units are not accepted SI units, and are in decreasing use. In Sections 1.4–1.6, the quantities of illuminance, intensity, and luminance, respectively, are discussed in more detail.

As with radiometric terms, spectral photometric quantities describe the distribution of the quantity as a function of the wavelength. By integrating the spectral luminous quantity over wavelength with a desired filter function, one finds the total luminous quantity over the desired spectral range

$$f_v = \int_0^{\infty} h(\lambda) f_v(\lambda) d\lambda. \quad (1.2)$$

Conversion between radiometric and photometric units is accomplished by taking into account the response of the CIE standard observer. The functional form is given by

$$f_v(\lambda) = K(\lambda) f_e(\lambda), \quad (1.3)$$

TABLE 1.3 Alternate Units for Illuminance and Luminance

Unit	Abbreviation	Form
Illuminance		
Foot-candle	fc	lm/ft^2
Phot	ph	lm/cm^2
Luminance		
Apostilb	asb	$\text{cd}/\pi/\text{m}^2$
Foot-lambert	fL	$\text{cd}/\pi/\text{ft}^2$
Lambert	L	$\text{cd}/\pi/\text{cm}^2$
Stilb	sb	cd/cm^2

where $f_v(\lambda)$ is the spectral photometric quantity of interest, $f_e(\lambda)$ is the analogous spectral radiometric term, and $K(\lambda)$ is the luminous efficacy, which is a function of wavelength, λ , and has units of lm/W. The luminous efficacy describes the CIE observer response to visible electromagnetic radiation as a function of wavelength. The profile of $K(\lambda)$ is dependent on the illumination level, because of the differing response of the eye's detectors. For example, for light-adapted vision, that is, photopic vision, the peak in the luminous efficacy occurs at 555 nm. For dark-adapted vision, that is, scotopic vision, the peak in the luminous efficacy is at 507 nm.* Equation (1.3) is often rewritten as

$$f_v(\lambda) = CV(\lambda)f_e(\lambda), \quad (1.4)$$

where $V(\lambda)$ is the luminous efficiency,[†] which is a unitless quantity with a range of values between 0 and 1, inclusive, and C is a constant dependent on the lighting conditions.[‡] For photopic vision, $C = C_p = 683$ lm/W, and for scotopic vision $C = C_s = 1700$ lm/W. Note that the constant for photopic conditions is in agreement with the definition of the candela as discussed previously. The difference between the two lighting states is due to the response of the cones, which are not saturated for typical light-adapted conditions, and rods, which are saturated for light-adapted conditions. The lumen is realistically defined only for photopic conditions, so for scotopic cases the term “dark lumen” or “scotopic lumen” should be used. Light levels between scotopic and photopic vision are called mesopic, and are comprised of a combination of these two states. Figure 1.2 shows the luminous efficiency of the standard observer for the two limiting lighting conditions as a function of wavelength. Note that while similar, the response curves do not have the same shape. Light-adapted vision is broader than dark-adapted vision.

To calculate the luminous quantity when the spectral radiant distribution is known, one must integrate over wavelength using the luminous efficacy and the filter function as weighting terms. Using Equations (1.1) and (1.3), we arrive at

$$f_v = \int_0^{\infty} h(\lambda)K(\lambda)f_e(\lambda)d\lambda. \quad (1.5)$$

where the limits of integration are set to 0 and ∞ , since the functional forms of $K(\lambda)$ and $h(\lambda)$ take into account the lack of response outside the visible spectrum and the wavelength range(s) of interest, respectively. Alternatively, one can determine the radiometric quantity with knowledge of the analogous photometric spectral distribution; however, only the value over the visible spectrum is calculated.

* For photopic vision, the luminous efficacy may be labeled as $K(\lambda)$ or $K_p(\lambda)$, and for scotopic vision, the luminous efficacy may be labeled as $K'(\lambda)$ or $K_s(\lambda)$. In Equation (1.3), $K(\lambda)$ is used in generic form to indicate either photopic or scotopic conditions.

[†] For photopic vision, the luminous efficiency may be labeled as $V(\lambda)$ or $V_p(\lambda)$, and for scotopic vision, the luminous efficiency may be labeled as $V'(\lambda)$ or $V_s(\lambda)$. In Equation (1.4), $V(\lambda)$ is used in generic form to indicate either photopic or scotopic conditions.

[‡] The constant C may be labeled as C or C_p for photopic vision and as C' or C_s for scotopic vision. In Equation (1.4), C is used in generic form to indicate either photopic or scotopic conditions.

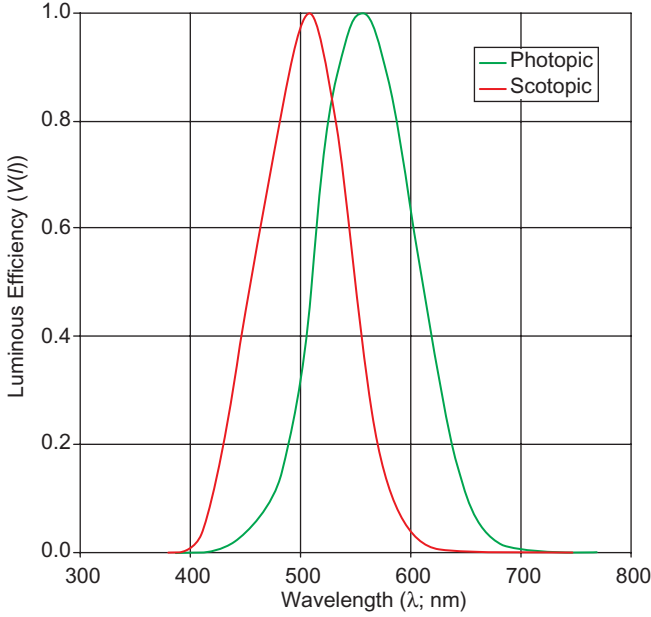


Figure 1.2 Luminous efficiency for photopic and scotopic conditions.

1.4 INTENSITY

Luminous or radiant intensity describes the distribution of light as a function of angle—more specifically a solid angle. The solid angle, $d\Omega$, in units of steradians (sr) is expressed by a cone with its vertex at the center of a sphere of radius r . As shown in Figure 1.3, this cone subtends an area dA on the surface of the sphere. For this definition, the solid angle is given by

$$d\Omega = \frac{dA}{r^2} = \sin\theta d\theta d\phi, \quad (1.6)$$

where, in reference to Figure 1.3, θ is the polar angle labeled as θ_0 and ϕ is the azimuthal angle around the central dotted line segment. By integrating the right-hand side of Equation (1.6) over a right-circular cone, one finds

$$\Omega = 4\pi \sin^2 \frac{\theta_0}{2}, \quad (1.7)$$

where θ_0 is the cone half angle as shown in Figure 1.3. A cone that subtends the entire sphere (i.e., $\theta_0 = \pi$) has a solid angle of 4π and one that subtends a hemisphere (i.e., $\theta_0 = \pi/2$) has a solid angle of 2π . Intensity is a quantity that cannot be directly measured via a detection setup, since detectors measure flux (or energy). A direct conversion to flux density is found by dividing through by the area of the detector while assuming the distribution is constant over the detector surface, but intensity

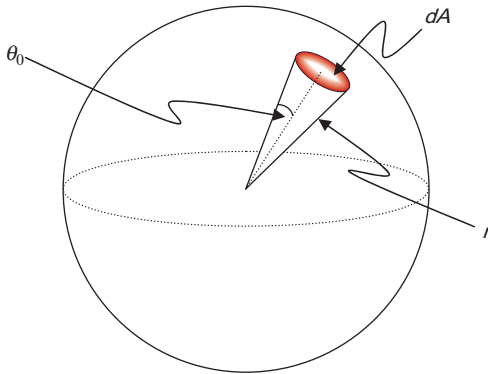


Figure 1.3 Geometry of sphere that defines solid angle.

requires knowledge of the angular subtense of the detector at the distance r from the source point of interest. Additionally, one must account for the overlap of radiation for an extended source when measurements are made in the near field. In the far field, where caustics are negligible, the intensity distribution can be inferred immediately from the flux density.

1.5 ILLUMINANCE AND IRRADIANCE

Illuminance and irradiance are the photometric and radiometric quantities respectively for the surface flux density on a target.* The material presented in this section is also applicable to the radiant and luminous exitances. These terms describe the spatial distribution of light since they integrate the luminance or radiance over the angular component. Detectors operate in this mode in the sense that the power incident on the detector surface is dependent on its area.† Figure 1.4 depicts the measurement of the flux density at a distance of r from a uniform point source by a detector of differential area dA . The subtense angle, θ , is between the normal to the detector area and the centroid from the source to the projected area of the detector area. The latter is simply the line segment joining the source (S) and the center of the detector (T), ST. Using Equation (1.6), the detector area subtends an elemental solid angle from the point S of

$$d\Omega = \frac{dA_{\text{proj}}}{r^2} = \frac{dA \cos \theta}{r^2}, \quad (1.8)$$

* Use of the terms irradiance or illuminance denote the generic surface flux density term, except when units are specifically listed. Henceforth, no subscripts (i.e., “e” or “v”) are used on any illumination quantities in order to denote the generic nature of the equations.

† Thus, the reason that a number of detectors have an active area of 1 cm^2 . The measured power then matches the flux density with the assumption that the flux is constant over the area of the detector.

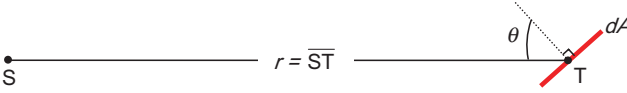


Figure 1.4 Measurement of the irradiance on a target area of dA at a distance r from a point source of flux output $d\Phi$. The target is oriented at θ with respect to the line segment joining points S and T.

where the differential projected area, dA_{proj} , will be discussed in detail in Section 1.6, Equation (1.10). Note that for the detector oriented at $\pi/2$ with respect to the line segment $\overline{ST}$, the elemental solid angle is 0. Conversely, for an orientation along the line joining the two entities, the elemental solid angle is at a maximum since the cosine term is equal to 1.

To find, for example, the irradiance due to a point source, we substitute for dA in its expression from Table 1.1 with that from Equation (1.8). The resulting equation contains an expression for the intensity, $I = d\Phi/d\Omega$, which can be substituted for,

$$E = \frac{d\Phi}{dA} = \frac{d\Phi \cos \theta}{r^2 d\Omega} = \frac{I \cos \theta}{r^2}. \quad (1.9)$$

Thus, for a point source, the flux density incident on the target decreases as a function of the inverse of the distance squared between the two objects, which is known as the inverse-square law. The cosine factor denotes the orientation of the target with respect to the line segment $\overline{ST}$.

Unlike experimental measurement of the intensity, it is easy to measure the flux areal density. With knowledge of the detector area and the power measured, one has determined the flux density at this point in space. This detection process integrates over all angles, but not all detectors measure uniformly over all angles, especially due to Fresnel reflections at the detector interface. Thus, special detector equipment, called a cosine corrector, is often included in the detection scheme to compensate for such phenomena. By measuring the flux density at a number of points in space, such as over a plane, the flux density distribution is determined. Additionally, unlike the intensity distribution, which remains constant with the distance from the optical system, the irradiance distribution on a plane orthogonal to the nominal propagation direction evolves as the separation between the optical system and plane is changed. Simply said, near the optical system (or even extended source), the rays are crossing each other such that local spatial (i.e., irradiance) distribution evolves with distance. In the far field, where the crossing of rays (e.g., caustics) are negligible, the irradiance distribution has the form of the intensity distribution. Thus, one is in the far field when these two distributions have little difference between their shapes. This point is discussed in more detail in Chapter 7.

1.6 LUMINANCE AND RADIANCE

Luminance and radiance are fundamental quantities of any emitter. As is shown in the next chapter, these terms are conserved in a lossless system, which drives the

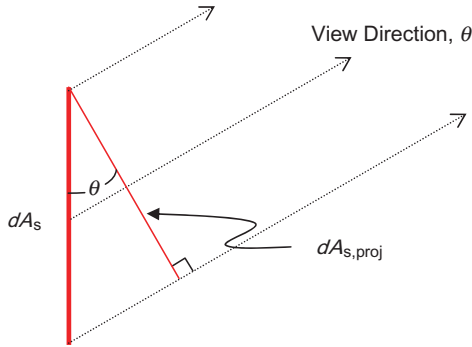


Figure 1.5 Depiction of $dA_{s,proj}$ of source element dA_s along a view direction θ .

ultimate limit of the performance of system design. Another term often used for luminance or radiance is brightness; however, within the vision community, brightness assumes the response for an actual observer to a prescribed area of an object* [10]. The radiance distribution[†] from a source describes the emission from each point on its surface as a function of angle. Thus, knowing the radiance distribution, one can ascertain the propagation of radiation through a known optical system. The result is that the intensity and flux density distributions, and for that matter, the radiance distribution can be determined at arbitrary locations in the optical system. In conclusion, radiance provides the best quantity to drive the design process of an illumination system for two reasons:

- An accurate source model implies that an accurate model of an illumination system can be done. The process of developing an accurate source model is a focus of the next chapter.
- Radiance is conserved, which provides the limit on system performance while also providing a comparison of the performances of different systems. The implications of conservation of radiance are presented in detail in Chapter 2.

The radiance distribution is measured as a function of source flux per projected unit area (i.e., spatial considerations), $dA_{s,proj}$, per unit solid angle (i.e., angular considerations), $d\Omega$. The projected area takes into account the orientation of the target surface with respect to the source orientation, which is shown in Figure 1.5.[‡] Pursuant to this figure, the elemental projected area is given by

$$dA_{s,proj} = dA_s \cos \theta, \quad (1.10)$$

* Though the difference between brightness and luminance is minor, caution indicates that using the accepted terms of luminance and radiance is preferred.

[†] For simplicity, I use radiance to mean both luminance and radiance.

[‡] Note that some express the radiance as the flux per unit area per projected solid angle. While the latter definition does not agree with the standard, it does provide for the same quantity.

where dA_s is the actual area of the source and θ is the view angle. Substituting Equation (1.10) into the expression for the radiance

$$L = \frac{d^2\Phi}{\cos\theta dA_s d\Omega}. \quad (1.11)$$

In connection with radiance, there are two standard types of distributions: Lambertian and isotropic. These radiance distributions are described in more detail in the next two subsections.

1.6.1 Lambertian

When a source or scatterer is said to be Lambertian, it implies its emission profile does not depend on direction,

$$L(\mathbf{r}, \theta, \phi) = L(\mathbf{r}) = L_s, \quad (1.12)$$

where $\mathbf{r}$ is the vector describing the points on the surface, θ is the polar emission angle, and ϕ is the azimuthal emission angle. Thus, only the dependence on the areal projection modifies the radiance distribution. Substituting Equation (1.12) into Equation (1.11) while using Equation (1.6) for the definition of the elemental solid angle and then integrating over angular space, one obtains

$$\int \frac{d^2\Phi}{dA_s} = \int dM = \int_{4\pi} L_s \cos\theta d\Omega. \quad (1.13)$$

The solution of this gives

$$M = L_s \int_0^{2\pi} \int_0^{\pi/2} \cos\theta \sin\theta d\theta d\phi \quad (1.14)$$

$$= M_{\text{lam}} = \pi L_s,$$

where the “lam” subscript denotes a Lambertian source. The exitance for a Lambertian source or scatterer is always the product of π and its radiance, L_s . Similarly, one can integrate over the area to determine the intensity

$$\int \frac{d^2\Phi}{d\Omega} = \int dI = \int_D L_s \cos\theta dA.$$

Integrating provides

$$\begin{aligned} I &= \cos\theta \int_D L_s dA \\ &= I_{\text{lam}} = I_s \cos\theta, \end{aligned} \quad (1.15)$$

where D is the surface of the source and

$$I_s = \int_D L_s dA. \quad (1.16)$$

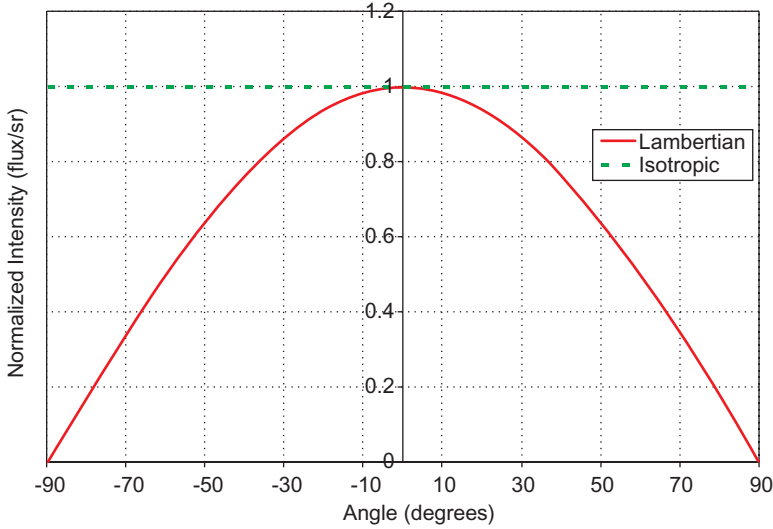


Figure 1.6 Plot of normalized intensity profiles for Lambertian and isotropic spatially uniform planar sources.

Thus, the intensity for a Lambertian source changes as a function of the cosine of the view angle. For a spatially uniform planar emitter of area A_s , the intensity distribution is

$$I_{lam} = A_s L_s \cos \theta. \quad (1.17)$$

Figure 1.6 shows the normalized intensity profile for an emitter described by Equation (1.17). Note that as the angle is increased with respect to the normal of the surface (i.e., $\theta = 0$ degrees), less radiation is seen due to the projection characteristics. In other words, as the polar angle increases, the projected area decreases, resulting in a reduction in intensity. It is an important point that while Lambertian means independent of direction, it does not preclude the reduction in intensity due to the projection of the emission area.

1.6.2 Isotropic

An isotropic source or scatterer is also said to be uniform, but such a source accounts for its projection characteristics. Thus, the cosine dependencies shown in Equations (1.15) and (1.17) are compensated within the radiance function by

$$L(\mathbf{r}, \theta, \phi) = \frac{L(\mathbf{r})}{\cos \theta} = \frac{L_s}{\cos \theta}. \quad (1.18)$$

The exitance and intensity for such a source are

$$M_{iso} = 2\pi L_s, \quad (1.19)$$

and

$$I_{\text{iso}} = I_s, \quad (1.20)$$

where the “iso” subscripts denote isotropic and I_s is as defined in Equation (1.16). For the spatially uniform planar emitter of area A_s , the intensity distribution is given by

$$I_{\text{iso}} = A_s L_s. \quad (1.21)$$

Therefore, an isotropic source has twice the exitance compared with an analogous Lambertian source. It also has a constant intensity profile as a function of angle for a planar emitter. The normalized isotropic intensity profile of Equation (1.21) is also shown in Figure 1.6. Realistically, only point sources can provide isotropic emission. There are no independent sources that provide isotropic illumination, but the integration of tailored optics and an emitter can provide such.

1.7 IMPORTANT FACTORS IN ILLUMINATION DESIGN

The transfer of light from the source to the target typically has two important parameters: transfer efficiency and the distribution at the target. Transfer efficiency is of particular importance due to increasing needs of energy efficiency—that is, electricity costs are increasing, and concerns over environmental effects are rising. However, often counter to efficiency requirements is the agreement of the achieved distribution to the desired one. For example, the desire for uniformity can easily be achieved by locating a bare source a great distance from the target plane, albeit at the expense of efficiency. Therefore, in the case of uniformity, there is a direct trade between these two criteria such that the illumination designer must develop methods to provide both through careful design of the optical system. Lesser criteria include color, volume requirements, and fabrication cost. As an example, the design of a number of illumination optics is driven by the reduction of cost though the process time to manufacture the overall system. So not only must the designer contend with trades between efficiency and uniformity, but also cost constraints and other important parameters. The end result is a system that can be quickly and cheaply fabricated with little compromise on efficiency and uniformity. The selection and functional form of the important criteria for an illumination design provide a merit function that is used to compare one system with another. The development of a merit function based on the material presented here is discussed in detail in Chapter 7.

1.7.1 Transfer Efficiency

Transfer efficiency, η_t , is defined as the ratio of flux at the target, Φ_{target} , to that at the input, which is most often the flux emitted from the source, Φ_{source} :

$$\eta_t = \frac{\Phi_{\text{target}}}{\Phi_{\text{source}}}. \quad (1.22)$$

This simple definition includes all emission points, all angles, and the entire spectrum from the source. Criteria based on spatial positions, angular domain, and spectrum can be used to select limited ranges of the emission from the source. The spectral radiance, $L_{e,\lambda}$, or spectral luminance, $L_{v,\lambda}$, describe terms* that replace the flux components in Equation (1.22),

$$\eta_t = \frac{\iiint h_{\text{target}}(\mathbf{r}, \Omega, \lambda) L_{\text{target},\lambda}(\mathbf{r}, \Omega, \lambda) d\mathbf{r} d\Omega d\lambda}{\iiint h_{\text{source}}(\mathbf{r}, \Omega, \lambda) L_{\text{source},\lambda}(\mathbf{r}, \Omega, \lambda) d\mathbf{r} d\Omega d\lambda}, \quad (1.23)$$

where h_{target} and h_{source} are filters denoting the functional form for position ($\mathbf{r}$, which has three spatial components such as x , y , and z), angle (Ω , which is the solid angle and composed of two components such as θ , the polar angle, and ϕ , the azimuthal angle), and spectrum (λ) for the target and source, respectively. $L_{\text{target},\lambda}$ and $L_{\text{source},\lambda}$ denote the spectral radiance or luminance for the target and source, respectively. Note that h_{target} and h_{source} almost always have the same form for the spectrum, but the position and angular aspects can be different. These filter and radiance functions can be quite complex analytically, thus they are often approximated with experimental measurements or numerical calculations. The end result is that solving Equation (1.23) can be rather cumbersome for realistic sources and target requirements, but it provides the basis of a driving term in illumination design, étendue, and the required conservation of this term. It is shown in Chapter 2 that the étendue is related to radiance, thus Equation (1.23) expresses a fundamental limit and thus efficiency for the transfer of radiation to the target from the source.

1.7.2 Uniformity of Illumination Distribution

The uniformity of the illumination distribution defines how the modeled or measured distribution agrees with the objective distribution. The illumination distribution is measured in at least one of three quantities:

- Irradiance or illuminance: measured in flux/unit area
- Intensity: measured in flux/steradian, or
- Radiance or luminance: measured in flux/steradian/unit area.

For the radiometric quantities (i.e., irradiance, radiant intensity, and radiance), the unit of flux is the watt (see Table 1.1). For the photometric quantities (i.e., illuminance, luminous intensity, and luminance) the unit of flux is the lumen (see Table 1.2). These quantities are described in depth in Sections 1.3–1.6. Other quantities have been used to express the illumination distribution, but they are typically hybrids or combinations of the three terms provided above.

Uniformity is determined by comparing the sampled measurement or model with the analogous one of the goal distribution. Note that the irradiance/illuminance and intensity distributions can be displayed with two orthogonal axes, while the

* The spectral radiance is used for radiometric calculations, while the spectral luminance is used for photometric calculations.

radiance and luminance quantities require a series of depictions to show the distribution. There are a multitude of methods to determine the uniformity, including:

- Peak-to-valley variation of the distribution, Δf :

$$\Delta f = \max[f_{\text{model}}(i, j) - f_{\text{goal}}(i, j)] - \min[f_{\text{model}}(i, j) - f_{\text{goal}}(i, j)]. \quad (1.24)$$

- Variance of the distribution compared with the goal, σ^2 :

$$\sigma^2 = \frac{1}{(mn-1)} \sum_{j=1}^n \sum_{i=1}^m [f_{\text{model}}(i, j) - f_{\text{goal}}(i, j)]^2. \quad (1.25)$$

- Standard deviation of the distribution, σ : the square root of the variance provided in Equation (1.25).

where the f terms denote the selected quantity that defines uniformity (e.g., irradiance or intensity). The model and goal subscripts denote the sampled model and goal distributions, respectively. The terms i and j are the counters for the m by n samples, respectively, over the two orthogonal axes. The term root mean square (RMS) deviation is found by taking the square root of an analogous form of the bias-corrected variance expressed in Equation (1.25). For the RMS variation, σ_{rms} , the factor in front of the summation signs in Equation (1.25) is replaced with $1/nm$.^{*} In all cases, a value of 0 for the uniformity term (e.g., Δf or σ^2) means that the uniformity, or agreement with the target distribution, is perfect for the selected sampling. The choice of the uniformity metric is dependent on the application, accuracy of the modeling, and ease of calculation. The RMS variation of the selected quantity is the standard method of calculating the uniformity of a distribution.

1.8 STANDARD OPTICS USED IN ILLUMINATION ENGINEERING

There is a multitude of types of optics that can be used in the design of an illumination system. These types can essentially be broken down into five categories: refractive optics (e.g., lenses), reflective optics (e.g., mirrors), total internal reflection (TIR) optics (e.g., lightpipes), scattering optics (e.g., diffusers), and hybrid optics (e.g., catadioptric Fresnel elements or LED pseudo collimators). In the next five subsections, I discuss each of these optic types in more detail. Each of these optic types has different utility in the field of illumination design, and this book uses separate chapters to discuss and delineate these. As a general rule of thumb, reflective optics provide the most “power” to spread light, but at the expense of tolerance demands and typical higher absorption losses. Refractive optics allow more compact systems to be built, but at the expense of dispersion, an increased number of elements, and higher fabrication costs due to alignment issues and postproduction manufacturing

^{*} See MathWorld, hosted and sponsored by Wolfram Research; specifically, see <http://mathworld.wolfram.com/StandardDeviation.html> for further discussion of this topic.

demands. TIR optics provide in theory the optimal choice, except when one includes potential leakage due to both scattering and a lack of fulfilling the critical-angle condition at all interfaces. Scatter is not typically employed as a method to provide a critical illumination distribution, but, rather, a means to make the distribution better match uniformity goals at the expense of efficiency. Scatter is also used for subjective criteria, such as look and feel of the optical system. Hybrid optics that employ reflection, refraction, scatter, and TIR are becoming more prevalent in designs since they tend to provide the best match in regards to optimizing goals.

1.8.1 Refractive Optics

Refractive optics are a standard tool in optical design, primarily used for imaging purposes, but they can also be used in illumination systems. There is a variety of refractive optics used in illumination systems. In fact, the types of refractive optics are too numerous to discuss here, but they range from standard imaging lenses to arrays of pillow optics to protective covers. Examples of refractive optics in use in illumination systems include: (i) singlet lens for projector headlamp, (ii) pillows lens array for transportation applications, (iii) nonimaging Fresnel lens for display purposes, and (iv) a protective lens for an automotive headlight. The first three examples are using the refraction to assist in obtaining the target illumination distribution, but (iv) has minimal, if not negligible, impact on the distribution of light at the target. The primary purpose of the latter lens is to protect the underlying source and reflector from damage due to the environment.

Image-forming, refractive optics are not optimal for illumination applications in the sense that they do not maximize concentration, which is defined in the next chapter. The reason for the limitation is due to aberrations as the f -number is decreased. This topic for imaging systems has been investigated in detail [11]. In Reference [11], it is pointed out that the theoretical best is the Luneberg lens, which is a radial gradient lens shaped as a perfect sphere and has a refractive index range from 1.0 to 2.0 inclusive. This lens is impossible to manufacture due to the index range spanning that of vacuum to high-index flint glass, such as Schott LaSF35. Realistic lenses, such as an $f/1$ photographic objective or an oil-immersed, microscope objective, are examples of imaging lenses that provide the best performance from a concentration point of view. These two examples do not provide optimal concentration, and they also suffer in that their sizes are small, or in other words, while their $f/\#$ might be small, accordingly the focal length and thus diameter of the clear aperture are also small. However, there are cases when nonoptimal imaging, refractive systems are preferred over those that provide optimal concentration. An example is a pillow lens array to provide homogeneity over a target. Each of the pillow lenses creates an aberrated image of the source distribution on the target plane, thus the overlap of these distributions due to the lens array can be used to create an effective uniform distribution. Additionally, nonimaging lenses are an option, but this type of lens almost always involves other phenomena, especially TIR, to accomplish their goals. These optics are treated in Section 1.8.5 on hybrid optics. Refractive, nonimaging lenses use, in part, imaging principles to great effect to transfer the light from the source to the target. Two examples are the nonimaging

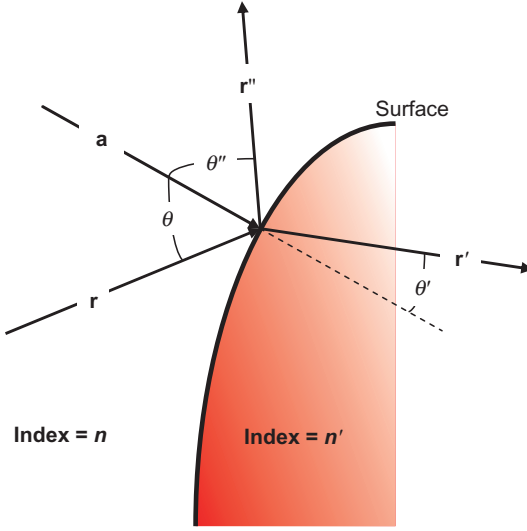


Figure 1.7 Schematic of refraction, reflection, and total internal reflection at an optical surface. The incident ray vector is shown by $\mathbf{r}$ in index n . The refracted ray vector is $\mathbf{r}'$ in index n' . The reflected and TIR ray vectors are $\mathbf{r}''$ in index n . For TIR to occur $n > n'$, and $\theta \geq \theta_c$.

Fresnel lens, which has been discussed in the literature [12], and the catadioptric lens used to collimate as best as possible the output from an LED. The latter is a focus of Chapter 7 on optimization and tolerancing of nonimaging optics. Nonimaging, refractive designs are making great use of tailoring, where tailoring is defined as the designation of the optical surface to provide a prescribed distribution at the target in order to meet design goals. Tailoring is discussed further in the next chapter and the succeeding chapters in this book. The systems and principles presented here and others are described in more depth in Chapters 4–6.

Refractive optics use materials of differing indices of refraction, n , to alter the propagation path of light from the source to the target. Refraction, displayed in Figure 1.7, is governed by the law of refraction, often called Snell's law,

$$n \sin \theta = n' \sin \theta', \quad (1.26)$$

where θ and θ' are the angles of incidence and refraction, respectively, and n and n' are the indices of refraction in source (i.e., object) and target (i.e., image) spaces, respectively. In other words, the unprimed notation denotes the quantities prior to refraction, and the primed notation indicates the quantities after refraction. In illumination systems, three-dimensional (3D) ray tracing is needed, thus the vector form is preferred,

$$n \mathbf{r} \times \mathbf{a} = n' \mathbf{r}' \times \mathbf{a}, \quad (1.27)$$

where $\mathbf{a}$ is the surface normal into the surface as shown in Figure 1.7, and $\mathbf{r}$ and $\mathbf{r}'$ are the unit vectors depicting the rays. Equation (1.27) also implies that $\mathbf{r}$, $\mathbf{r}'$, and $\mathbf{a}$

are coplanar, which is a necessary condition for Snell's law. However, this equation is not tractable to ray tracing, thus a different form is employed. First, find the cross product of the surface normal $\mathbf{a}$ with the two sides of Equation (1.27) to give after rearrangement of the terms

$$n'\mathbf{r}' = n\mathbf{r} + (n'\mathbf{a} \cdot \mathbf{r}' - n\mathbf{a} \cdot \mathbf{r})\mathbf{a}. \quad (1.28)$$

Next, we note that $\mathbf{a} \cdot \mathbf{r} = \cos \theta$ and $\mathbf{a} \cdot \mathbf{r}' = \cos \theta'$, which upon substitution gives

$$n'\mathbf{r}' = n\mathbf{r} + (n'\cos \theta' - n\cos \theta)\mathbf{a}. \quad (1.29)$$

Finally, this expression is rewritten in component form with the use of the direction cosines $\mathbf{r} = (L, M, N)$ in incident space, $\mathbf{r}' = (L', M', N')$ in refraction space, and $\mathbf{a} = (a_L, a_M, a_N)$. In component form, Snell's law for ray-tracing purposes is written

$$\begin{aligned} n'L' &= nL + (n'\cos \theta' - n\cos \theta)a_L, \\ n'M' &= nM + (n'\cos \theta' - n\cos \theta)a_M, \text{ and} \\ n'N' &= nN + (n'\cos \theta' - n\cos \theta)a_N. \end{aligned} \quad (1.30)$$

Note that the refraction angle θ' must be found prior to the determination of the direction cosines after refraction. Snell's law as given in Equation (1.26) is used for this purpose

$$\theta' = \arcsin\left(\frac{n}{n'}\sin \theta\right). \quad (1.31)$$

Substitution of Equation (1.31) into Equation (1.30) and diving through by n' provides the ray path after refraction for the purposes of numerical ray tracing. The only caveat to the implementation of Equation (1.30) is that $\mathbf{r} = (L, M, N)$, $\mathbf{r}' = (L', M', N')$, and $\mathbf{a} = (a_L, a_M, a_N)$ are all unit vectors.

1.8.2 Reflective Optics

Reflective optics are also a standard tool in optical design, but admittedly imaging design uses them less than refractive ones. In illumination design, reflective optics have been more prevalent than refractive ones. The primary reasoning is that it is easier to design the optics that provide optimal concentration between the input and output apertures, and the tolerance demands for nonimaging systems are less in comparison to imaging ones. There is a wealth of reflective optics in use in illumination designs from conic reflectors, such as spherical, parabolic, elliptical, and hyperbolic, to edge-ray designs (EDs) [13], such as the compound parabolic concentrator (CPC), to tailored edge-ray designs (TEDs) [13, 14]. The conic reflectors typically involve imaging requirements, and illumination designs use these properties to capture the light emitted by a source. Nonimaging designs use the edge-ray principle to transfer optimally the light from source to target. EDs are optimal in two dimensions (2D), called troughs, but 3D designs, called wells, do not transfer some skew rays from the source to target. ED assumes a constant acceptance angle with uniform angular input, but TED uses a functional acceptance angle. Thus, TED

accommodates realistic sources while providing tailoring of the distribution at the target dependent on requirements and the source characteristics. Systems employing the edge-ray principle are based on their parent, conic designs, thus there are two classes: elliptical and hyperbolic.

Tailored designs are currently at the forefront of reflector design technology in the nonimaging optics sector. Note that tailored designs started within the realm of reflective illumination optics, but they are now present in all sectors of nonimaging optics, from refractive to hybrid. Tailored reflectors use one of two methods: discrete faceted or continuous/freeform reflectors. The former is comprised of pseudo-independent, reflective segments to transfer the light flux. The freeform design can be thought of as a faceted design but made up of an infinite number of individual segments. The designs are typically smooth, although discontinuities in the form of cusps (i.e., the first derivative is discontinuous) are sometimes employed, but steps are not allowed since that is the feature of faceted designs. While tailored designs employ optics principles in order to define their shape, the chief goal is to maintain a functional acceptance angle profile. Recently, applications that are not based entirely on the optics of the problem have dictated that the edge-ray can limit performance of the overall system. These systems use interior rays to motivate the reflector development, thus this domain is called nonedge-ray reflector design (NERD) [15]. One application is optical pumping of gain material, which introduces some imaging properties in order to obtain more efficient laser output at the expense of transfer efficiency [16].

Examples of reflective illumination designs include: (i) luminaire for room lighting, (ii) faceted headlight, (iii) faceted reflector coupled to a source for projection and lighting applications (i.e., MR16), and (iv) a reflector coupled to an arc source for emergency warning lighting. For luminaires for architectural lighting, the shape of the reflector is either a conic or an arbitrary freeform surface. In both of these cases, there is minimal investment in the design of the reflector, but rather a subjective look and feel is the goal of the design process. In recent years, luminaire designs for specific architectural applications, such as wall-wash illumination, has integrated tailoring in order to provide better uniformity and a sharp cutoff in the angular distribution. Increasingly, especially with the development of solid-state lighting, luminaire reflectors employ some level of tailoring in order to meet the goals of uniformity while also maintaining high transfer efficiency. They must also meet marketing requirements that drive their appearance in both the lit and unlit states. For example, the headlight shown in (ii) must adhere to stringent governmental illumination standards (e.g., ECE [European] or FMVSS/SAE [US]), but must also conform to the shape of the car body and provide a novel appearance.

Specular reflectors employ either reflective materials, such as polished aluminum or silver; reflective materials deposited on substrates; or dielectric thin film coating stacks deposited on polished substrates. For reflective deposition, the coatings are typically aluminum, chromium, or even silver and gold. The substrate is typically an injection-molded plastic. For dielectric coatings, the substrates can be absorbing, reflective, or transmissive. Dielectric coatings on reflective substrates protect the underlying material but also can be used to enhance the overall reflectivity. Dielectric coatings placed on absorbing or transmissive substrates can be used

to break the incident spectrum into two components so that the unwanted light is removed from the system. An example is a hot mirror that uses a dichroic placed on an absorbing substrate. The visible light is reflected by the coating, while the infrared (IR) light is absorbed by the substrate. In all cases, the law of reflection can be used to explain the behavior of the reflected rays. For reflection, which is also shown in Figure 1.7, the incident ray ($\mathbf{r}$) angle, θ , is equal to that of the reflected ray ($\mathbf{r}''$) angle, θ'' ,

$$|\theta| = |\theta''|. \quad (1.32)$$

Note that we update the prime notation used for refraction to a double prime notation to denote reflection. The development employed in the previous section can be used to determine the direction cosine equations for the propagation of a reflected ray by using the law of refraction with $n' = n'' = -n$ and $\mathbf{r}' = -\mathbf{r}''$.^{*} Using this formalism, Snell's law for reflection gives $\theta = -\theta''$, and Equations (1.28) and (1.29) upon reflection give

$$\mathbf{r}'' = \mathbf{r} - (\mathbf{a} \cdot \mathbf{r}'' + \mathbf{a} \cdot \mathbf{r})\mathbf{a} = \mathbf{r} - 2(\mathbf{a} \cdot \mathbf{r})\mathbf{a}, \quad (1.33)$$

and

$$\mathbf{r}'' = \mathbf{r} - 2\mathbf{a} \cos \theta. \quad (1.34)$$

The direction cosines are then written,

$$\begin{aligned} L'' &= L - 2a_L \cos \theta \\ M'' &= M - 2a_M \cos \theta, \text{ and} \\ N'' &= N - 2a_N \cos \theta. \end{aligned} \quad (1.35)$$

Thus, Equations (1.30) and (1.31) are essentially the same, and often optical ray-tracing software implements Equation (1.20) with the caveat that a negative index of refraction is used with an antithetical reflection vector.

1.8.3 TIR Optics

TIR optics use the “frustration” of refraction in order to propagate light within the higher index material that traps the light. Examples of TIR illumination optics include (i) large-core plastic optical fibers, (ii) lightpipes, (iii) lightguides for display applications, and (iv) brightness enhancement film. TIR is not a standard in imaging design, used sparingly in components like prisms and fibers. In illumination optics, TIR optics are gaining in popularity, especially in hybrid form as discussed in Section 1.8.5. This popularity gain is due to real and perceived benefits, including higher transfer efficiency, assistance with homogenization, compact volume, and the component provides guiding along its length. Unfortunately, to turn these benefits into realistic TIR optics, it requires a sizable investment in design and fabrication.

^{*} A number of authors say that reflection is a special case of refraction, but they only point out that $n' = n'' = -n$, while neglecting to point out that $\mathbf{r}' = -\mathbf{r}''$.

The higher transfer efficiency is due to the 100% Fresnel reflection as long as the critical angle, θ_c , or greater, is satisfied. The critical angle is governed by Snell's law

$$\theta_c = \arcsin\left(\frac{n'}{n}\right), \quad (1.36)$$

where this angle describes the point at which light that would refract from the higher index material (n) into a lower-index material (n') at an angle of $\theta' = 90^\circ$ to the surface normal (i.e., along the tangent to the surface). With TIR, the light will be trapped within the optic, propagating along the reflection ray vector $\mathbf{r}''$, as shown in Figure 1.7. The only perceived losses are at the input and output apertures of the TIR medium due to Fresnel reflections. With angles of incidence in the range of $\pm 45^\circ$, the Fresnel reflection, typically considered a loss, is around 4% per uncoated interface. However, with only the two surfaces, input and output, a transfer efficiency of greater than 90% is still difficult to obtain because of scatter from both volume and surface imperfections that frustrate the TIR condition, leakage due to lack of holding to the critical angle condition, and back reflection from other surfaces of the optic due to shape and guiding requirements for the TIR optic. For example, some lightpipes used in automotive, dashboard applications have been known to have transfer efficiency around 10% to the output.

Homogenization requires careful development of the TIR optic to ensure proper mixing of the source(s) within the medium. A standard rule-of-thumb is that to obtain homogenization, the length of the TIR optic must be ten times or greater than its width. This length-to-width ratio thus demands a rather large volume in comparison to non-TIR methods, such as pillow lens arrays. Additionally, input and output locations within the optic may make increased demands on the system design, resulting in a trade on other benefits. These trade studies are often driven by volume limitations, thus while the final design can remain compact, one must give up uniformity and efficiency.

Notwithstanding, TIR optics are often the best choice for the design of illumination systems, especially those requiring remote or distributed lighting. The most prevalent TIR optics in the illumination sector are lightpipes and lightguides. A standard use of lightpipes is for discrete lighting applications, such as indicators and buttons, while lightguides are used for distributed lighting seen, for example, in backlit displays. This distinction is arbitrary, thus the terms lightpipe and lightguide are often used interchangeably. There are three aspects to the design of the lightpipes: the coupling of the source to it, the efficient propagation of the captured flux, and the outcoupling with the desired pattern. The incoupling usually involves secondary optics, proximity, and widening of the optic after the input aperture. The design of the guide after coupling, as stated previously, is driven by volume limitations, output demands, and reduction of leakage. The design of the lightpipe propagation sections can be tedious and complex, requiring intense user interaction due to the limited number of algorithms. The outcoupling is achieved through four standard methods: refraction at the output surface; narrowing of the guide optic; bumps, holes, or scatter sites on the surface of the lightpipe; and shaping of the optic to allow controlled leakage. These topics and others are considered in detail in Chapter 6 of this book.

1.8.4 Scattering Optics

Scatter optics purposely use surface roughness or volume scattering in order to approach the desired distribution. They do such by trading off transfer efficiency because of scattering outside the desired distribution or even backscattering. Additionally, scatter centers placed on or within optical components provide a method to lessen hot spots (i.e., regions resulting in high illumination levels) or cold spots (i.e., regions resulting in low illumination levels); to reduce or smooth the gradient between neighboring illuminated regions, especially since human observation is drawn to the gradient rather than the level; to provide a desired appearance, lit and/or unlit; frustrate TIR; and provide mixing to hide individual sources, especially those from different spectral sources. Scatter is also unintentionally present in all illumination optics because nothing can be made perfectly. Surfaces will never be completely smooth, and volumes will have some level of inclusions that cause scattering.

Scattering sites are used in all types of components: reflective, refractive, TIR, and absorbing/opaque. Examples include: (i) a tailored diffuser, (ii) a cosine-corrected diffuser, (iii) a roughened housing for a transmissive, automotive optic, and (iv) a rough, metallic luminaire. The first two examples use scatter to provide a distinct distribution of light, where the first is using surface scattering and the second is using volume scattering. For a number of illumination optics incorporating roughened surfaces, the scatter is placed on the optic to not only assist in meeting required standards, but to provide a smoother appearance to the illumination distribution while also hiding to some extent the interior operation of the illumination system. For example, a metal luminaire reflector incorporates a nonspecular surface to provide a fairly uniform and soft illumination of the lit environment.

Surface scattering, both intentional and unintentional, is explained by the bidirectional surface distribution function (BSDF). The BSDF is the generic form to describe the scatter distribution, but the bidirectional reflectance distribution function (BRDF) and bidirectional transmittance distribution function (BTDF) denote the reflection and transmission distributions, respectively. A special case of the BSDF for weakly scattering surfaces, such as those found in optical systems, is the Harvey model. These terms describe how the light is scattered as a function of incident and observation angles. Volume scattering is described by model distributions, such as Mie scatter, Henyey–Greenstein [17], and the Gegenbauer model [18]. Finally, there is Lambertian and isotropic (uniform) scatter for both surfaces and volumes. Both provide equal scatter as a function of angle, but isotropic scatter compensates for the cosine falloff due to the projection of the scattering surface. The reader is directed to the book *Optical Scattering* by Stover for more information about scattering theory and measurement [19].

1.8.5 Hybrid Optics

Hybrid illumination optics, also called combination optics, are those that integrate by design at least two of the four optical phenomena described in the previous

sections. With this definition, one could construe that a source coupled to a reflector in close proximity to a lightguide is a hybrid optic; however, this is an example of a system with distinct optical components each doing a separate function. Rather, the definition for hybrid illumination optic is further refined by demanding that the optical phenomena must be integrated into a single optical component. Note that a simple lightpipe that has planar surfaces at the input and output apertures performs two refractions; however, these refractions are just incidental, and the optic is considered solely a TIR structure. Alternatively, if the input or output surfaces were shaped to provide a desired distribution, then the optic would be a hybrid since it integrates by design two of the four optical phenomena previously presented.

Examples of hybrid illumination optics include (i) a catadioptric Fresnel element, (ii) a backlighting lightguide using scatter sites for output, and (iii) an integrated coupler-illuminator for LED sources. In (i), the light first uses refraction at the entrance aperture of the Fresnel element, followed by TIR at the next intercept, and then refraction at the output aperture. For (ii), the lightguide uses TIR to propagate the light along its length and then scatter to eject the light from the lightguide. The integrated coupler-illuminator for LED sources is the leading example of the utility of hybrid illumination optics. This optic integrates the source coupling and then the illumination into the desired target distribution. This method ensures compactness, higher efficiency, and required uniformity. Design algorithms are being proposed for the development of these hybrid optics. One in particular, the simultaneous multiple surfaces (SMS) method, can incorporate reflection, TIR, and refraction into the design of a single optic to provide desired wave fronts upon emission [20]. The SMS design method is described in detail in Chapter 4. The ray propagation algorithms explained in the previous sections are used to explain the behavior of the individual ray paths, but the final design of the optic is based on the combination of the many ray paths. For all of these reasons and others, hybrid illumination optics are at the forefront of illumination engineering, and will likely prove for the near future to be the selected optic for demanding illumination requirements. Hybrid illumination optics are discussed in more depth in Chapter 7.

1.9 THE PROCESS OF ILLUMINATION SYSTEM DESIGN

There are many steps that go into the design and fabrication of an optical system, in particular an illumination system. Figure 1.8 shows a flowchart for the steps of the process. The primary steps are divided into two phases: design and fabrication. The solid arrows inline with the steps shows the natural progression of the steps, while the exterior solid lines show potential iteration paths in order to ensure that the built system meets requirements. The path on the left-hand side of the primary path shows the potential iterations involved in the design phase, while the path on the right-hand side of the primary path shows the potential iterations initiated in the fabrication phase. The dashed lines show optional paths to circumvent a step in the primary path. A description of each step is provided here:

Design Phase

- *Concept.* The concept for the system is determined, including the type of optics that can potentially be used, goals, and a description of the system.
- *Baseline.* The design specifications, including efficiency, distribution, color, cost, and volume. Additionally, an étendue analysis is done in order to determine what source is required to meet the goals. Note that it is acceptable practice to set design specifications higher than required in order to alleviate tolerance and fabrication concerns.
- *Literature.* The designer investigates the literature taking into account the goals. Sometimes, the type of optics to be employed can be determined during this step, for example, reflector optics coupled to a source. This step can essentially be ignored if the designer has experience with the application being pursued.
- *Initial Study.* This step is also called conceptual study. If the type of optics has not already been determined, this step studies the effectiveness of various system configurations. At the end of this step, the chosen path and configuration for future design is selected.
- *Design.* The illumination optics are developed. In this step, the important optimization and tolerance parameters are determined. At the end of this step, the untoleranced system is at a minimum close to meeting design goals.
- *Optimization.* An optimization or perturbation routine is employed to improve the performance of the system. The optimization parameters are determined during the previous design steps. At the end of this step, the untoleranced system exceeds design specifications.

Fabrication Phase

- *Tolerance.* The important tolerance parameters are determined during the design steps. Two types of tolerancing are done: parameter sensitivity and Monte Carlo analysis. For parameter sensitivity, each of the important parameters is perturbed over a prescribed range to ascertain system performance. For Monte Carlo analysis, all of the important parameters are perturbed randomly in order to map out the performance of realistic systems. Unfortunately, this step is often ignored by many designers due to time constraints and a perception that the fabricated system will easily meet requirements. However, this neglect often leads to the major reason that manufactured illumination systems do not meet specifications.
- *Fabrication.* The components of the illumination system are integrated into a CAD layout for the complete system, and then the individual components are manufactured.
- *Testing.* There are two potential aspects to this step: component testing and system testing. For component testing, which is an optional substep, individual parts of the complete illumination system are tested to ensure they conform to the design specification. The parts are then integrated into the systems.

Finally, the system is tested. Comparison between the measurements and design results ensures the system meets requirements.

Note that the design iteration path only includes steps within the design phase, but the fabrication iteration path includes steps within both phases. The reasoning is that during the fabrication phase, it may be determined that the base design is not suitable for the application due to tolerance or fabrication demands. In this event, the design phase may have to be reopened to obtain a system that is able to meet the goals.

Figure 1.8 provides a fairly complete description of the steps involved in the successful completion of an illumination system; however, certain designs may require less or more steps. Ultimately, this determination is based on the complexity of the system, experience of the designer, level of system requirements, and evolving

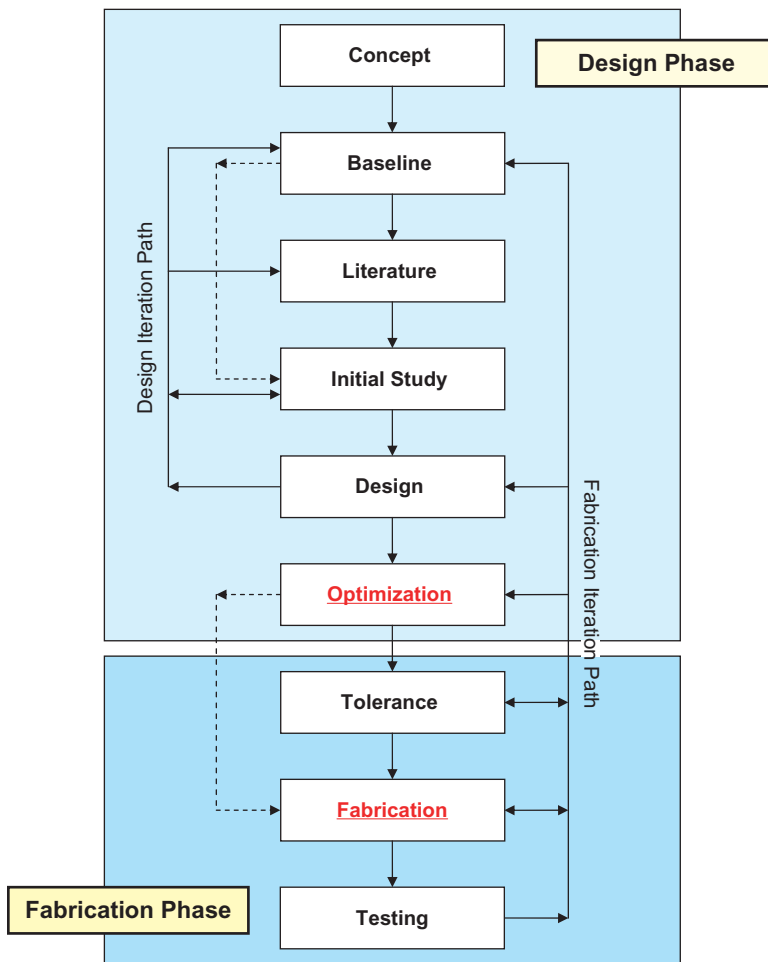


Figure 1.8 Flowchart showing the phases of the design process of an illumination system. The Optimization and Fabrication steps indicate a path that many designers do at their peril.

metrics and goals. Additionally, in order to address fabrication realities, a novel method of integrating design, optimization, and tolerancing into a single large step is employed. Essentially, this implies three phases of the system development process: concept phase, design phase, and fabrication phase. The concept phase would include the concept, baseline, literature, and initial study steps.

Now we introduce a number of terms, especially the radiometric and photometric units, in order to establish a basis for illumination optics design, analysis, and measurement. The understanding of the units is imperative in order to make more effective illumination system designs.

1.10 IS ILLUMINATION ENGINEERING HARD?

As presented in Sections 1.1, 1.8, and 1.9, there are a number of factors that comprise illumination design: from the original concept to the design, including optimization and tolerancing. Followed by the design work is the fabrication and integration into the final system. As stated the primary goals are to illuminate with the desired distribution while maximizing the transfer efficiency to the target. In Section 1.7, I commented upon other factors that also drive the design process, including appearance, color, and cost. Often, these additional factors actually drive the design, but it is difficult to quantify their effect on the design. In other words, it is hard to include them in a merit function that describes the performance of the system. Thus, the designer typically uses an iterative process for the flowchart shown in Figure 1.8. For this one reason, illumination design, though mistakenly perceived as straightforward, is a time-consuming (some may at times say tedious!) process. The final system meets demands (which are often relaxed during the design process) in order to obtain an allowable yield in the production process.

Additionally, other factors that slow down the design process are the limited number of design algorithms, the lack of formally trained people in the field of illumination design, the lack of connection between the design and fabrication industries, and requirements that violate conservation laws, especially étendue. The lack of algorithms is due to the stage of this industry—the field of energy efficient illumination systems that meet distribution demands is in its infancy compared with other optical design fields. A number of these algorithms are not mathematically intensive, but, rather, based on intuition and heuristic results. The methods are also specific to the application, such as lightpipes or luminaires, so their utility is limited. Additionally, the methods often treat sparingly the tolerance and manufacturing issues. In time, the realities of the production will be integrated into these models, but at this time, iteration is the best solution. Also, effective illumination design deals with extended sources deployed in nonsequential ray-tracing systems. The tracing of individual rays is approximately a linear process (requiring the paraxial approximation for refractive optics) in the sense that we can follow the path of a ray through the system, but the superposition of all the rays provides a rather nonlinear result. In other words, the illumination designer has to contend with caustics that make the process that much more difficult. The lack of formally trained individuals

in the field is primarily due to the infancy and perception of the field. The perception is based on the world around us in which lighting is seen to be a commodity and thus inexpensive. However, a sizable amount of energy is wasted in the lighting industry, from unwanted light such as glare, inefficient sources, and solutions that provide look rather than form. As energy costs continue to rise and an increasing need to tailor illumination, more individuals will be formally trained in the field of illumination design. All industries contend with a disconnect between the designers and the fabricators. The illumination industry is especially susceptible to this due to the lack of formally trained engineers, the infancy of the industry, and the lack of terminology to describe tolerance and fabrication demands. Continued investment in the connection between these two arms of the illumination industry will reduce the system design costs. Finally, every illumination designer must at one time in their career try to violate the conservation of étendue, which is explained in the next chapter. Étendue and its conservation is the driving physics in the design of illumination systems, and by attempting to violate it, the designer will actually learn its properties and consequences. The increasing demands on illumination systems mean that we are often approaching the limit of étendue, but note that it is hard to calculate the étendue limit for a system because of the realistic emission characteristics of the source. Also, it is difficult to convey the concept of conservation of étendue to those not trained in the field of illumination design. One goal of this book is to assist the new illumination designer to the concept of étendue and its limitations while also providing some information to educating the untrained, especially your boss or customer!

I end this section with the observation that it does not fail to surprise me the amount of time in the optical design phase that must be dedicated to the appearance of the system, the volume limitations, and the yield requirements. On every project, I learn something new that increases my experience in the field of illumination design. So, at this time, the field of illumination design tends to be a rather hard discipline requiring the user to gain experience through trial and error, intuition, and the growing amount of literature in the field.

1.11 FORMAT FOR SUCCEEDING CHAPTERS

In Chapter 2, the important illumination topics of étendue, its conservation, concentration, acceptance angle, and skew invariant are introduced. In Chapter 3, the concept of étendue is expanded by discussing the “squeezing” of it. Étendue squeezing is a technique to address dilution and mixing within the optical design process. In Chapter 4, the SMS method is presented. Examples are presented to illustrate this design method. Chapter 5 looks at the application of nonimaging optics in the field of solar energy concentrators. Novel methods including those of Chapter 4 are used to design efficient, limited tracking solar concentrators. Chapter 6 gives a detailed treatment of lightpipe design, analysis, and modeling methods. Chapter 7 provides information about the successful use of optimization and tolerancing methods in the design of nonimaging optics.

REFERENCES

1. R.J. Koshel, Why illumination engineering? in *Nonimaging Optics and Efficient Illumination Systems IV, SPIE Proc.* 6670, 667002 (2007). Bellingham, WA.
2. CIE, *Commission Internationale de l'Éclairage Proceedings 1924*, Cambridge University Press, Cambridge, UK (1926).
3. CIE, *Bureau Central de la CIE* **1**, Section 4, **3**, 37 (1951).
4. W. Crummett and A.B. Western, *University Physics: Models and Applications*, Wm. C. Brown Publishers, Dubuque, IA (1994).
5. R.W. Boyd, *Radiometry and the Detection of Optical Radiation*, John Wiley and Sons, New York (1983).
6. E. Hecht, *Optics*, 4th ed., Addison Wesley, Reading, MA (2001).
7. R. Winston, J.C. Miñano, and P. Benítez, *Nonimaging Optics*, Elsevier Academic Press, Burlington, MA (2005).
8. J.M. Palmer, Getting Intense about Intensity, *Optics and Photonics News* **6**(2), 6 (1995) or see the website <http://www.optics.arizona.edu/Palmer/intenopn.html> (last accessed 9 September 2012).
9. NIST, *The International System of Units (SI)*, NIST Special Publication 330, 2001 ed., B.N. Taylor, ed., NIST, Gaithersburg, MD (2001). <http://physics.nist.gov/Pubs/pdf.html> or <http://physics.nist.gov/cuu/Units/bibliography.html> for downloadable version of this and other NIST publications in the area of units and measurements.
10. D. Malacara, *Color Vision and Colorimetry: Theory and Applications*, SPIE, Bellingham, WA (2002).
11. W.T. Welford and R. Winston, *High Collection Nonimaging Optics*, Academic Press, Inc., San Diego, CA (1989).
12. R. Leutz and A. Suzuki, *Nonimaging Fresnel Lenses: Design and Performance of Solar Concentrators*, Springer, Berlin (2001).
13. R. Winston, J.C. Miñano, and P. Benítez, *Nonimaging Optics*, Elsevier Academic Press, Burlington, MA (2005).
14. J. Chaves, *Introduction to Nonimaging Optics*, CRC Press, Boca Raton, FL (2008).
15. R.J. Koshel, Non-edge ray reflector design (NERD) for illumination systems, *Novel Optical Systems Designs and Optimization II, SPIE Proc.* 4092, (2000), p. 71. Bellingham, WA.
16. R.J. Koshel and I.A. Walmsley, Non-edge ray design: improved optical pumping of lasers, *Opt. Eng.* **43**, 1511 (2004).
17. S.L. Jacques and L.-H. Wang, Monte Carlo modeling of light transport in tissues, in A.J. Welch and M.J.C. van Gemert, eds., *Optical Thermal Response of Laser Irradiated Tissue*, Plenum Press, New York (1995), pp. 73–100.
18. A.N. Yaroslavsky, I.V. Yaroslavsky, T. Goldbach, and H.-J. Schwarzaier, Influence of the scattering phase function approximation on the optical properties of blood determined from the integrating sphere measurements, *J. Biomed. Opt.* **4**, 47–53 (1999).
19. J.C. Stover, *Optical Scattering*, 2nd ed., SPIE Press, Bellingham, WA (1995).
20. P. Benítez, J.C. Miñano, J. Blen, R. Mohedano, J. Chaves, O. Dross, M. Hernández, and W. Falicoff, Simultaneous multiple surface optical design method in three dimensions, *Opt. Eng.* **43**, 1489–1502 (2004).