

TITLE

Carmela Filosa

Cover art:
Photography:

A catalogue record is available from the Eindhoven University of Technology Library

ISBN: 978-90-386-3972-7

Copyright © 2017 by C. Filosa.

All rights are reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of the author.

title

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een
commissie aangewezen door het College voor
Promoties, in het openbaar te verdedigen

door

Carmela Filosa

geboren te Torre del greco, Italië

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter: prof.dr.

1^e promotor: prof.dr. W.L. IJzerman

copromotor: dr. J.H.M. ten Thije Boonkkamp

leden:

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Contents

1	Introduction	3
1.1	Motivation	3
1.2	Methods and results	3
1.3	Content of this thesis	3
2	Illumination optics	5
2.1	Radiometric and photometric variables	5
2.2	Reflection and refraction law	10
2.3	Fresnel's equations	11
3	Ray tracing	19
3.1	Ray tracing for two-dimensional optical systems	19
3.2	Monte Carlo ray tracing	20
3.3	Quasi-Monte Carlo method	35
4	Ray tracing on phase space	37
4.1	Phase space concept	37
4.2	The edge-ray principle	37
4.3	Phase space ray tracing	37
5	Two different approaches to compute the boundaries in target phase space	39
5.1	The α -shapes approach	39
5.2	The two-faceted cup	42
5.3	Results for a TIR collimator	42
5.4	The triangulation refinement approach	42
5.5	The two-faceted cup	42
5.6	Results for a TIR collimator	42
5.7	Results for a Parabolic reflector	42
5.8	Results for the Compound Parabolic Concentrator (CPC)	42
6	The inverse ray mapping method: analytic approach	43
6.1	Explanation of the method	43
6.2	The two-faceted cup	43

6.3	The multi faceted cup	43
6.4	Results for the two-faceted cup	43
6.5	Results for the multi-faceted cup	43
6.6	Discussions	43
7	The extended ray mapping method	45
7.1	Explanation of the method	45
7.2	Bisection procedure	45
7.3	Results for a parabolic reflector	45
7.4	Results for two different kind of TIR-collimators	45
8	Extended ray mapping method to systems with Fresnel reflection	47
9	Discussion and conclusions	49
	List of publications	53
	Curriculum Vitae	55
	Acknowledgments	57
	Bibliography	59

List of symbols

$d\Omega$	Solid angle
θ	Angle between the direction of the solid angle and the normal ν
θ_i	Angle between the incident ray and the normal ν
θ_r	Angle between the reflected ray and the normal ν
θ_t	Angle between the transmitted ray and the normal ν
ϕ	Angle that the ray located on line i forms with respect to the optical axis

Chapter 1

Introduction

1.1 Motivation

1.2 Methods and results

1.3 Content of this thesis

Chapter 2

Illumination optics

This chapter provides some concepts of illumination optics used in this thesis. We start explaining the difference between radiometry and photometry. In particular, we focus on the photometric variables, defining them both in three and two dimensions. The reflection and refraction laws and the phenomenon of total internal reflection are explained. The last paragraph of the chapter gives a brief introduction to Fresnel reflection.

2.1 Radiometric and photometric variables

Radiometry is concerned with the measurement of electromagnetic radiation across the entire electromagnetic spectrum. Photometry is the subfield of radiometry that takes into account only the portion of the electromagnetic spectrum corresponding to the visible light, [1]. Radiometry deals with radiometric quantities. An important radiometric quantity is the radiant flux Φ_r (unit watt, [W]) which is the total energy emitted from a source or received by a target per unit time:

$$\Phi_r = \frac{dQ}{dT}, \quad (2.1.1)$$

where Q is the energy and T the time.

In illumination optics the measurement of light is given in terms of the impression that it gives on the human eye. Therefore, illumination optics deals with photometric variables. The most important photometric variables are defined as follows using the same notation adopted by Chaves in [2]. The luminous flux Φ (unit lumen, [lm]) is defined as the perceived power of light by the human eye. The radiant and the luminous flux are related by the luminous efficacy function, unit [lm/W], which tells us how many lumen there are for each Watt of power at a given wavelength. The luminous efficacy reaches its maximum at a wavelength of 555 nm where it is equal to 683 lm/W. We may normalize the luminous efficacy function with its maximum value of 683. This normalized function is the dimensionless luminosity function $\bar{y}(\lambda)$ shown in Figure 2.1 where λ is the wavelength.

The luminous flux corresponding to one Watt of radiation power at any wavelength is given by the product of 683 lm/W and the luminosity function at the same wavelength,

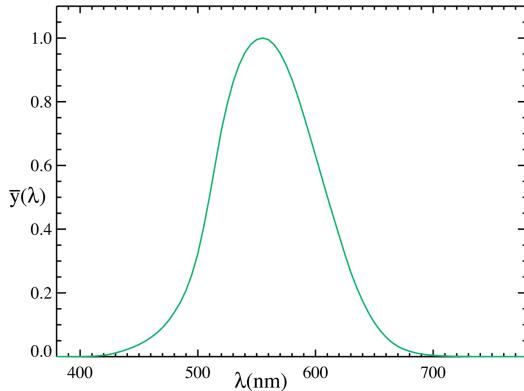


Figure 2.1: Luminosity function $\bar{y}(\lambda)$: relation between the eye's sensitivity and the wavelength of light. The luminosity function is dimensionless, [3].

i.e. $683 \bar{y}(\lambda)$. Hence, the total luminous flux Φ has unit lumen [lm] and it is defined as:

$$\Phi = 683 \int_0^{\infty} \Psi_r(\lambda) \bar{y}(\lambda) d\lambda \quad (2.1.2)$$

where $\Psi_r(\lambda)$ is the power in watt per unit wavelength (unit [W/m]).

A beam of light can be described as a collection of parallel light rays, where a light ray can be interpreted as a path along which the energy travels. The luminous flux $d\Phi$ incident on a surface is called illuminance E (unit [lm/m²]) and is defined as:

$$E = \frac{d\Phi}{dA}, \quad (2.1.3)$$

where dA is an infinitesimal area receiving radiation. The density of light emitted by a point source in a given direction is determined by the solid angle. The solid angle on a given direction is defined by the infinitesimal surface area dS of a sphere subtended by the radius of that sphere and by the rays emitted by the center on that direction, [4]. Indicating with r the radius of the sphere, the infinitesimal solid angle $d\Omega$ defined by dS is given by:

$$d\Omega = \frac{dS}{r^2}. \quad (2.1.4)$$

The solid angle on the entire sphere is $\Omega = 4\pi$, its unit is the steradian [sr] and it is usually defined on a unit sphere. The luminous intensity I (unit candela (cd), [cd = lm/sr]) is defined as the luminous flux $d\Phi$ per solid angle $d\Omega$ and is given by:

$$I = \frac{d\Phi}{d\Omega}. \quad (2.1.5)$$

The luminance L (unit [cd/m²]) is the luminous flux per unit solid angle $d\Omega$ and per unit projected area $\cos\theta dA$ where θ is the angle that the normal ν to the area dA makes with the direction of the solid angle $d\Omega$, as shown in Figure 2.2. L is given by:

$$L = \frac{d\Phi}{\cos\theta dA d\Omega}. \quad (2.1.6)$$

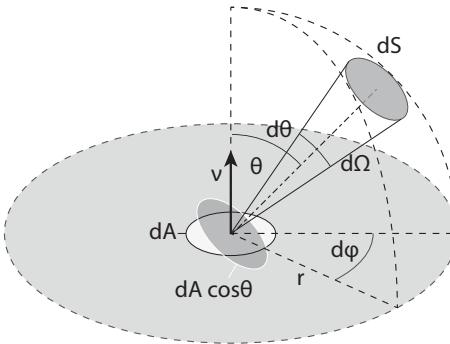


Figure 2.2: Solid angle $d\Omega$ in a direction making an angle θ with the normal to the area dA .

Note that from (2.1.5) and (2.1.6) we can derive a relation between the intensity and the luminance. The intensity I emitted by the infinitesimal area dA is given by:

$$I = \frac{d\Phi}{d\Omega} = L(\mathbf{x}, \theta) \cos \theta dA. \quad (2.1.7)$$

When the luminance is uniform over a finite area A , the luminous intensity emitted in the direction θ is equal to:

$$I(\theta) = L(\theta) A \cos \theta. \quad (2.1.8)$$

Thus, when $L(\mathbf{x}, \theta)$ does not depend on the position and the direction (i.e. $L(\mathbf{x}, \theta) = L$), we obtain Lambert's cosine law:

$$I(\theta) = I_0 \cos \theta. \quad (2.1.9)$$

where $I_0 = I(0) = LA$.

Finally, the étendue U (unit [m sr]) describes the ability of a source to emit light or the capability of an optical system to receive light, [5]. The quantity dU is defined as:

$$dU = n^2 \cos \theta dA d\Omega. \quad (2.1.10)$$

where n is the index of refraction of the medium in which the surface A is immersed. In optics the étendue is considered to be a volume in phase space (or an area for two-dimensional systems). This concept will be clarified in Chapter 4 in which we treat the phase space in more detail. An important property of the étendue is that it is conserved within an optical system in absence of absorption. We now show, using the approach of Chaves in [2], how conservation of this quantity can be derived. Consider a light ray emitted from an infinitesimal area dA_1 to the area dA_2 . Suppose that the centers of dA_1 and dA_2 are located at a distance d to each other, see Figure 2.3. Indicating with ν_1 and ν_2 the normals to the surfaces dA_1 and dA_2 , respectively and with θ_1 and θ_2 the angles that the central ray forms with ν_1 and ν_2 , respectively, the flux $d\Phi_1$ passing through dA_2 coming from dA_1 and the corresponding solid angle



Figure 2.3: dA_1 and dA_2 are two surfaces with normals ν_1 and ν_2 , respectively. Their centers are located at a distance d . θ_1 and θ_2 are the angles made by the central ray with the normals ν_1 and ν_2 , respectively.

$d\Omega_1$ are defined as:

$$\begin{aligned} d\Phi_1 &= L \cos \theta_1 dA_1 d\Omega_1, \\ d\Omega_1 &= \frac{dA_2 \cos(\theta_2)}{d^2}. \end{aligned} \quad (2.1.11)$$

Similarly, the flux $d\Phi_2$ passing through dA_1 coming from dA_2 is equal to:

$$\begin{aligned} d\Phi_2 &= L \cos \theta_2 dA_2 d\Omega_2 \\ d\Omega_2 &= \frac{dA_1 \cos \theta_1}{d^2}. \end{aligned} \quad (2.1.12)$$

Then from Eq. (2.1.10) we obtain the following relations:

$$\begin{aligned} dU_1 &= n^2 dA_1 \cos \theta_1 d\Omega_1 = \frac{n^2 dA_1 \cos \theta_1 dA_2 \cos \theta_2}{d^2}, \\ dU_2 &= n^2 dA_2 \cos \theta_2 d\Omega_2 = \frac{n^2 dA_2 \cos \theta_2 dA_1 \cos \theta_1}{d^2} \end{aligned} \quad (2.1.13)$$

for dA_1 and dA_2 , respectively. From the previous equations we can conclude that $dU_1 = dU_2$ and therefore the étendue dU is conserved along a beam of light. Since also the flux through the areas dA_1 and dA_2 is conserved, the following relation holds:

$$L := n^2 \frac{d\Phi}{dU} = \text{constant}. \quad (2.1.14)$$

In the optical systems we will consider in this work, the source and the target are located in the same medium (air) with $n = 1$, so the luminance L equals the basic

luminance $L^* = L/n$ at the source and the target of the system.

In this thesis we consider two-dimensional optical systems. Hence, the definitions of the photometric parameters have to be given in two dimensions. An infinitesimal line segment of length da that emits a light beam and the ray that makes an angle θ with the normal ν are considered, see Fig. 2.4.

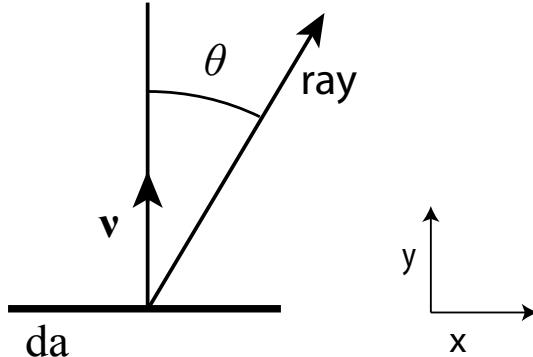


Figure 2.4: Ray emitted by an infinitesimal line segment da that makes an angle θ with respect to the line normal ν .

The two-dimensional illuminance (unit [lm/m]) denotes the luminous flux falling on an infinitesimal line segment of length da and it is given by:

$$E = \frac{d\Phi}{da}. \quad (2.1.15)$$

The luminous intensity (unit [lm/rad]) is the luminous flux per angle $d\theta$:

$$I = \frac{d\Phi}{d\theta}. \quad (2.1.16)$$

The two-dimensional luminance (unit [$\text{lm}/(\text{rad m})$]) is given by:

$$L = \frac{d\Phi}{\cos \theta da d\theta}. \quad (2.1.17)$$

Thus the following relation holds:

$$I = L(x, \theta) \cos \theta da \quad (2.1.18)$$

where x is a certain position at the light source da . Finally, the étendue dU (unit [mrad]) in two dimensions is given by:

$$dU = n \cos \theta da d\theta. \quad (2.1.19)$$

In order to determine the light distribution on a surface and to compute the photometric variables on that surface, we need to understand how the light emitted from the source propagates. In the field of geometric optics the light propagation is described by light rays. The propagation of a light ray traveling through different media is determined by the reflection and refraction law. In the following we introduce these two laws and we explain the total internal reflection phenomenon.

2.2 Reflection and refraction law

A light ray is described by a position vector \mathbf{x} on a surface and a direction vector \mathbf{t} and can be parameterized by the arc length s . Light rays travel in a homogeneous medium along straight lines, once they hit a reflective surface their direction changes. Denoting with \mathbf{t}_i the direction of the incident ray and with $\mathbf{\nu}$ the unit normal to the surface at the location of incidence, the direction \mathbf{t}_r of the reflected ray is given by:

$$\mathbf{t}_r = \mathbf{t}_i - 2(\mathbf{t}_i \cdot \mathbf{\nu})\mathbf{\nu}, \quad (2.2.1)$$

where the vectors \mathbf{t}_i and $\mathbf{\nu}$ are unit vectors and $\mathbf{t}_i \cdot \mathbf{\nu}$ indicates the scalar product between \mathbf{t}_i and $\mathbf{\nu}$. From Eq. (2.2.1) it follows that the vector \mathbf{t}_r is a unit vector too, indeed considering the scalar product $(\mathbf{t}_r, \mathbf{t}_r)$ we conclude:

$$\mathbf{t}_r \cdot \mathbf{t}_r = \mathbf{t}_i \cdot \mathbf{t}_i - 4(\mathbf{t}_i \cdot \mathbf{\nu})(\mathbf{t}_i \cdot \mathbf{\nu}) + 4(\mathbf{t}_i \cdot \mathbf{\nu})^2(\mathbf{\nu} \cdot \mathbf{\nu}) = 1. \quad (2.2.2)$$

The vectors \mathbf{t}_i , \mathbf{t}_r and $\mathbf{\nu}$ live all in the same plane. Defining the incident angle θ_i and the reflective angle θ_r such that $\theta_i, \theta_r \in [0, \pi/2]$. the reflection law states that $\theta_i = \theta_r$, see Fig. 2.5.

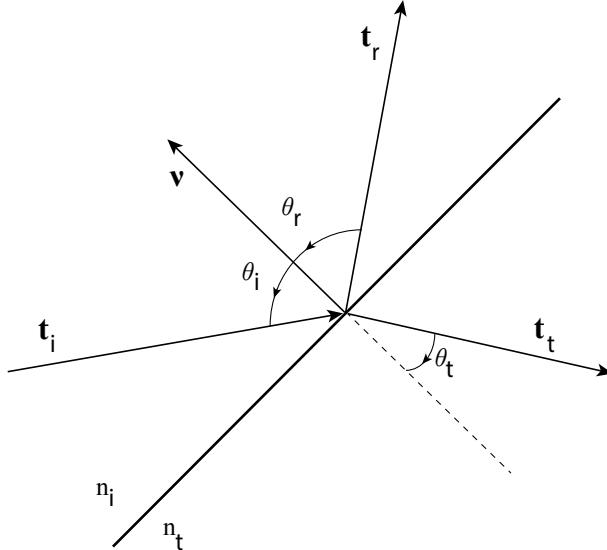


Figure 2.5: Propagation of a ray through two different media with index of refraction n_i and n_t .

When a ray propagates through two different media, its direction changes according to the law of refraction. Indicating with n_i the index of refraction of the medium in which the incident ray travels and with n_t the index of refraction of the medium of the transmitted ray, the direction \mathbf{t}_t of the transmitted ray is given by:

$$\mathbf{t}_t = n_{i,t} \mathbf{t}_i + \left[\sqrt{1 - n_{i,t}^2 + n_{i,t}^2(\mathbf{\nu} \cdot \mathbf{t}_i)^2} - n_{i,t}(\mathbf{\nu} \cdot \mathbf{t}_i) \right] \mathbf{\nu}, \quad (2.2.3)$$

where $n_{i,t} = n_i/n_t$, [2]. Note that in Eq. (2.2.1) the direction of the normal ν to the surface is not relevant for the computation of the direction of the reflective ray, since:

$$\mathbf{t}_r = \mathbf{t}_i - 2(\mathbf{t}_i \cdot \nu)\nu = \mathbf{t}_i - 2(\mathbf{t}_i \cdot -\nu)(-\nu), \quad (2.2.4)$$

however, this is not the case for Eq. (2.2.3), therefore in the latter case we need to specify the direction of ν which is usually chosen in such a way that the angle that it forms with the incident ray \mathbf{t}_i is smaller than or equal to $\pi/2$. Hence, if $(\mathbf{t}_i, \nu) \leq 0$ the normal ν directed inside the same medium in which travels the incident ray is taken as in Fig. 2.5, otherwise the normal $-\nu$ directed inside the same medium in which the transmitted ray will travel has to be considered.

Eq. (2.2.3) is only valid for

$$1 - n_{i,t}^2 + n_{i,t}^2(\nu \cdot \mathbf{t}_i)^2 \geq 0 \quad (2.2.5)$$

which implies that

$$\frac{n_t}{n_i} \geq \sqrt{1 - (\nu \cdot \mathbf{t}_i)^2} \quad (2.2.6)$$

from which we obtain:

$$n_t \geq n_i \sin \theta_i. \quad (2.2.7)$$

The angle θ_c for which the equality holds is

$$\theta_c = \arcsin \left(\frac{n_t}{n_i} \right) \quad (2.2.8)$$

and it is called the critical angle, [2]. When the incident angle θ_i is exactly equal to the critical angle θ_c , the square root in Eq. (2.2.3) is zero and the inner product $(\mathbf{t}_t, \nu) = 0$, hence the transmitted ray propagates parallel to the refractive surface. When $\theta_i > \theta_c$ the light ray is no longer refracted but is only reflected by the surface. This phenomenon is called total internal reflection (TIR). When TIR occurs, 100% of light is reflected and there is no loss of energy. Therefore, optical systems designed such that rays are reflected by TIR are very efficient. Light that hits an ordinary refractive surface can be reflected and refracted. The energy that is reflected and refracted is determined by the Fresnel's coefficients. In the next paragraph an overview of the Fresnel coefficients is given.

2.3 Fresnel's equations

In order to derive Fresnel's equations we need to describe light as an electromagnetic wave. It is therefore useful to study the light propagation from the perspective of electromagnetic theory which gives information about the incident, reflected and transmitted radiant flux density that are denoted with E_i , E_r and E_t , respectively. Any component of the electric field \mathcal{E} can be written as

$$\mathcal{E}(\mathbf{x}, T) = \mathcal{E}_0(\mathbf{x}) e^{i(k \cdot \mathbf{x} - \omega T)} \quad (2.3.1)$$

where \mathbf{x} is the position vector and T is the time. The amplitude $\mathcal{E}_0(\mathbf{x})$ is constant in time and $\omega = \frac{ck}{n}$ is the value of the angular frequency with c the velocity of light

and n the index of refraction in which the wave is traveling, which is the ratio of the speed of light c in vacuum and the speed of light v in the material. Note that the angular frequency can be also written as $\omega = vk$, in particular when a wave travels in vacuum $n = 1$ and $\omega = ck$. The vector \mathbf{k} has the same direction of the wave and its absolute value $|\mathbf{k}| = k = \frac{2\pi}{\lambda}$ is the wave number in vacuum, with λ the wavelength. Similarly, the magnetic field has the form:

$$\mathbf{B}(\mathbf{x}, T) = \mathbf{B}_0(\mathbf{x}) e^{i(\mathbf{k} \cdot \mathbf{x} - \omega T)}. \quad (2.3.2)$$

Light can be seen as an electromagnetic wave, that is an oscillating electric field \mathbf{E} and an oscillating magnetic field \mathbf{B} which propagates always perpendicular to \mathbf{E} . The electric field oscillates perpendicular to the wave propagation. Light is said to be polarized if the direction of the electric field is well defined. When the electric field propagates in different directions we talk about unpolarized light. By convention, we refer to the light's polarization as the direction of the electric field \mathbf{E} , [6] with respect to the incident plane that is defined by the incident and reflected rays as is shown in Fig. 2.6.

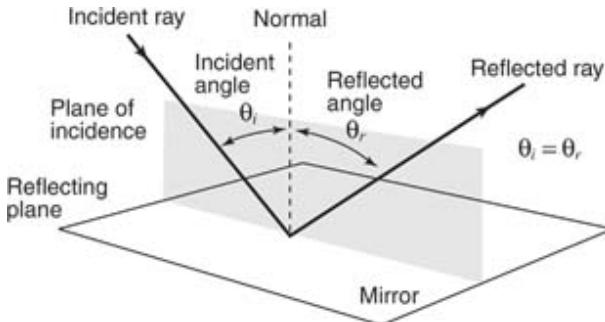


Figure 2.6: Light ray that hits a mirror located on the reflecting plane. The incident and the reflected ray leave in the same plane of the normal to the mirror that is called plane of incident.

In order to derive the Fresnel's coefficients the polarization of light must be taken into account. Those coefficients are obtained considering Maxwell's equations and the boundary conditions due to the conservation of energy. The details of Fresnel's equations are widely explained in the literature. In the following we provide Fresnel coefficients and we briefly explain their physical interpretation. We refer the reader to [7, 8] for more details. Fresnel's coefficients can also be derived using a different approach that does not involve Maxwell's equations, this method is explained in [9]. The following particular cases of light's polarization need are considered.

1. \mathbf{E} is perpendicular to the plane of incidence (see Fig. 2.7). In this case light is said to be *s*-polarized.
2. \mathbf{E} is parallel to the plane of incidence (see Fig. 2.8). In this case light is said to be *p*-polarized.

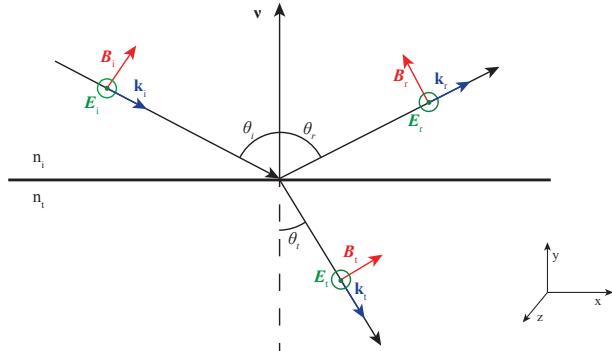


Figure 2.7: Propagation of an electromagnetic wave where \mathcal{E} is perpendicular to the incident plane. The components of \mathcal{E} are indicated with the green circles. The components of \mathcal{B} are indicated with red arrows.

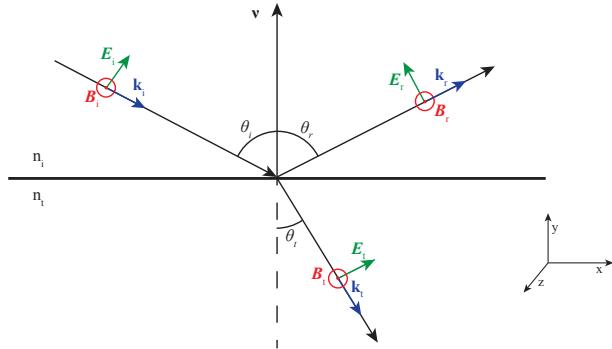


Figure 2.8: Propagation of an electromagnetic wave where \mathcal{E} is parallel to the incident plane. The components of \mathcal{B} are indicated with the red circle. The components of \mathcal{E} are indicated with green arrows.

Energy conservation gives the boundary conditions of the electromagnetic field at the plane of the interface (which is perpendicular to the incident plane). In the following we derive Fresnel's coefficients for case 1. Similarly, the Fresnel's coefficients can be derived for the second case.

For *s*-polarized light the tangential components of \mathcal{E} and \mathcal{B}/μ across the boundary between the two different media must be continuous. The continuity of the tangential component of \mathcal{E} leads to:

$$|\mathcal{E}_{0i}| + |\mathcal{E}_{0r}| = |\mathcal{E}_{0t}|, \quad (2.3.3)$$

while the continuity of the tangential component of \mathcal{B}/μ gives:

$$-\frac{|\mathcal{B}_i|}{\mu_i} \cos \theta_i + \frac{|\mathcal{B}_r|}{\mu_r} \cos \theta_r = -\frac{|\mathcal{B}_t|}{\mu_t} \cos \theta_t, \quad (2.3.4)$$

where the negative sign in front of $|\mathcal{B}_i|$ and \mathcal{B}_t is due to the convention that a positive direction is considered with increasing x . Since $\mathcal{B} = \mathcal{E}/v$, Eq. (2.3.4) can be written as

$$\frac{1}{\mu_i v_i} (|\mathcal{E}_i| - |\mathcal{E}_r|) \cos \theta_i = \frac{1}{\mu_t v_t} |\mathcal{E}_t| \cos \theta_t, \quad (2.3.5)$$

where we employed the fact that $v_i = v_r$, and $\theta_i = \theta_r$. Using Eq. (2.3.1) and $n = c/v$, the previous equation becomes:

$$\frac{n_i}{\mu_i} (|\mathcal{E}_{0i}| - |\mathcal{E}_{0r}|) \cos \theta_i = \frac{n_t}{\mu_i} |\mathcal{E}_{0t}| \cos \theta_t \quad (2.3.6)$$

Finally, assuming that $\mu_i = \mu_t = \mu_0$ and employing Eq. (2.3.3) we obtain:

$$\begin{aligned} r_s &= \frac{|\mathcal{E}_{0r}|_s}{|\mathcal{E}_{0i}|_s} = \frac{n_i \cos \theta_i - n_t \cos \theta_t}{n_i \cos \theta_i + n_t \cos \theta_t}, \\ t_s &= \frac{|\mathcal{E}_{0t}|_s}{|\mathcal{E}_{0i}|_s} = \frac{2n_i \cos \theta_i}{n_i \cos \theta_i + n_t \cos \theta_t}. \end{aligned} \quad (2.3.7)$$

The coefficients r_s and t_s are amplitude coefficients for the reflected and transmitted light. They are the perpendicular components of r and t for s -polarized light. Using Snell's law, that is $n_i \sin \theta_i = n_t \sin \theta_t$, the relations for r_s and t_s are simplified as follows:

$$\begin{aligned} r_s &= -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)}, \\ t_s &= -\frac{2 \sin \theta_t \cos \theta_i}{\sin(\theta_i + \theta_t)}. \end{aligned} \quad (2.3.8)$$

A similar argument for the p -polarized light leads to the calculation of the parallel components r_p and t_p of r and t . In case \mathcal{E} is parallel to the plane of incidence the amplitude coefficients are:

$$\begin{aligned} r_p &= \frac{n_t \cos \theta_i - n_i \cos \theta_t}{n_i \cos \theta_t + n_t \cos \theta_i}, \\ t_p &= \frac{2n_i \cos \theta_i}{n_i \cos \theta_t + n_t \cos \theta_i}, \end{aligned} \quad (2.3.9)$$

and their simplified relations are:

$$\begin{aligned} r_p &= \frac{\tan(\theta_i - \theta_t)}{\theta_i + \theta_t}, \\ t_p &= \frac{2 \sin \theta_t \cos \theta_i}{\sin(\theta_i + \theta_t) \cos(\theta_i - \theta_t)}. \end{aligned} \quad (2.3.10)$$

Furthermore, it can be checked that

$$\begin{aligned} t_s - r_s &= 1, \\ t_p + r_p &= 1. \end{aligned} \quad (2.3.11)$$

The amplitude coefficients are shown in Fig. 2.9 for the case in which light travels from a less dense to a more dense medium ($n_i < n_t$), that is external reflection. In

Fig. 2.10 the reflection coefficients are shown for the case in which $n_i > n_t$, that is internal reflection. Note from Fig. 2.9 that r_p approaches to 0 when θ_i approaches to θ_p and it gradually decreases reaching -1 for an incident angle $\theta_i = 90^\circ$. The angle θ_p is called Brewster's angle or polarization angle as only the component perpendicular to the incident plane is reflected at that angle and therefore light is perfectly polarized. Similarly, Fig. 2.10 shows that $r_p = 0$ for $\theta_i = \theta_{p'}$. It can be show that $\theta_p + \theta_{p'} = 90^\circ$. Both r_p and r_s reach 1 when $\theta_i = \theta_c$. θ_c is called the critical angle. Light that hits the incident plane with an incident angle equal to or greater than the critical angle is totally reflected back and no transmitted light is observed. This phenomenon is called total internal reflection.

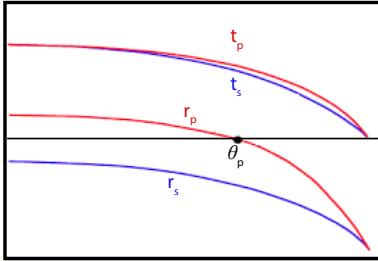


Figure 2.9: Amplitude coefficients of reflection and transmission as a function of the incident angle θ_i in the case of external reflection, i.e. $n_t < n_i$ ($n_t = 1$ and $n_i = 1.5$). θ_p is the polarization angle, [8].

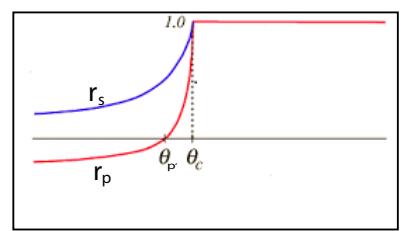


Figure 2.10: Reflection coefficients as a function of the incident angle θ_i in the case of internal reflection, i.e. $n_t > n_i$ ($n_t = 1.5$ and $n_i = 1$). θ_p' is the polarization angle and θ_c is the critical angle, [8].

The we introduce the Poynting vector \mathbf{P} that defines the energy flux of an electromagnetic field. It is measured in $[\text{W}/\text{m}^2]$, and it is given by:

$$\mathbf{P} = \frac{1}{\mu} (\mathbf{E} \times \mathbf{B}), \quad (2.3.12)$$

where $\mu = \frac{1}{\varepsilon v^2}$ is the permeability and ε the permittivity of the medium. In the following, the parameters for vacuum are indicated with the subscript 0. All quantities defined in the media of the incident, reflective and transmitted light are indicated with the subscripts i, r and t, respectively. Optical rays are perpendicular to the wave front of an electromagnetic wave and parallel to the Poynting vector, [10]. The irradiance E is defined as the average energy that crosses in unit time a unit area A perpendicular to the direction of the energy flow. Therefore, defining the average of the vector \mathbf{P} over the time as:

$$\langle \mathbf{P} \rangle_T = \frac{1}{T} \int_0^T \mathbf{P} dt \quad (2.3.13)$$

we can write the irradiance E as:

$$E = \langle \mathbf{P} \rangle_T = v \varepsilon |\mathbf{E}|^2. \quad (2.3.14)$$

Considering a beam of light that hits a surface such that an area A is illuminated, the incident, reflected and transmitted beams are $\mathbf{E}_i A \cos \theta_i$ $\mathbf{E}_r A \cos \theta_r$ and $\mathbf{E}_t A \cos \theta_t$,

respectively. The reflectance \mathcal{R} is the ratio of the reflected power to the incident power:

$$\mathcal{R} = \frac{|\mathbf{E}_r| \cos \theta_r}{|\mathbf{E}_i| \cos \theta_i} = \frac{|\mathbf{E}_{0r}|^2}{|\mathbf{E}_{0i}|^2} = r^2 \quad (2.3.15)$$

where the second equality holds because $v_i = v_t$, $\varepsilon_i = \varepsilon_t$ and $\theta_i = \theta_t$. Similarly, the transmittance \mathcal{T} is the ratio between the transmitted to the incident power:

$$\mathcal{T} = \frac{|\mathbf{E}_t| \cos \theta_t}{|\mathbf{E}_i| \cos \theta_i} = \frac{n_t \cos \theta_t}{n_t \cos \theta_i} \frac{|\mathbf{E}_{0t}|^2}{|\mathbf{E}_{0i}|^2} = \frac{n_t \cos \theta_t}{n_t \cos \theta_i} t^2. \quad (2.3.16)$$

Employing total energy conservation, that is:

$$\mathbf{E}_i A \cos \theta_i = \mathbf{E}_r A \cos \theta_r + \mathbf{E}_t A \cos \theta_t, \quad (2.3.17)$$

we can easily prove that:

$$\mathcal{R} + \mathcal{T} = 1. \quad (2.3.18)$$

The parallel and perpendicular components of \mathcal{R} and \mathcal{T} are:

$$\begin{aligned} \mathcal{R}_p &= r_p^2, \\ \mathcal{T}_p &= \frac{n_t \cos \theta_t}{n_t \cos \theta_i} t_p^2, \\ \mathcal{R}_s &= r_s^2, \\ \mathcal{T}_s &= \frac{n_t \cos \theta_t}{n_t \cos \theta_i} t_s^2. \end{aligned} \quad (2.3.19)$$

it can be show that

$$\begin{aligned} \mathcal{R}_s + \mathcal{R}_p &= 1, \\ \mathcal{T}_s + \mathcal{T}_p &= 1. \end{aligned} \quad (2.3.20)$$

For normal incidence, i.e. $\theta_i = 0$, there is no polarization and Eqs. (2.3.19) lead to:

$$\begin{aligned} \mathcal{R} &= \mathcal{R}_p = \mathcal{R}_s = \left(\frac{n_i - n_t}{n_t + n_i} \right)^2, \\ \mathcal{T} &= \mathcal{T}_p = \mathcal{T}_s = \frac{4n_i n_t}{(n_t + n_i)^2}. \end{aligned} \quad (2.3.21)$$

Many common light sources such as sunlight, halogen lighting, LED spotlights, and incandescent bulbs produce unpolarized light. In case of unpolarized light the amount of reflected and transmitted light is given by the average of reflectance \mathcal{R} and transmittance \mathcal{T} calculated considering first p -polarized light and then s -polarization, that is:

$$\begin{aligned} \mathcal{R} &= \frac{\mathcal{R}_p + \mathcal{R}_s}{2}, \\ \mathcal{T} &= \frac{\mathcal{T}_p + \mathcal{T}_s}{2}, \end{aligned} \quad (2.3.22)$$

where \mathcal{R}_p , \mathcal{R}_s , \mathcal{T}_p and \mathcal{T}_s are given in Eqs. (2.3.19).

With this overview we conclude this chapter. The notions given in Section 2.1 will be used in the entire thesis as our goal is to study the distribution of light at the target of some optical systems. In particular we will focus on the computation of the output intensity distribution. The reflection and refraction laws explained in Section 2.2 are needed to determine how the optical system changes the ray's direction every time that it hits a surfaces (or a line in the two-dimensional case). In Chapters 3, 4, 5, 6 and 7 only systems where the reflection and refraction laws play a role are considered. Systems with Fresnel reflection are treated in the last chapter. The amount of reflected and transmitted light is calculated using the Fresnel's equation (introduced in the last paragraph of this chapter). Since, we restrict ourselves to two-dimensional systems, the value of reflectance and transmittance will be computed using Eqs. (2.3.22).

Chapter 3

Ray tracing

Ray tracing is a geometric problem that describes the transport of light within optical systems. It uses single rays to describe the propagation of light through an optical system. The influence of diffraction on the transport of a ray is neglected and geometrical modeling of an optical system is considered. Generally, the method can be implemented for two or more dimensions and for any optical system. In the following a description of the ray tracing method in two dimensions is given.

3.1 Ray tracing for two-dimensional optical systems

Ray tracing is a procedure that can be applied to non-imaging systems. It consists of tracing each ray, which is considered to be a broken line, through the optical system. Given a Cartesian coordinate system (x, z) , a two-dimensional optical system symmetric with respect to the z -axis is defined. The optical system is formed by a source S , a target T and some optical components i where $i \in \{2, \dots, Nl - 1\}$ and Nl indicates the number of lines that form the system. S and T are indicated with the indexes 1 and Nl , respectively. The index of refraction of the medium in which line i is located is indicated with n_i . Every ray emitted by S (line 1) can hit some optical components $i \in \{2, \dots, Nl - 1\}$ before reaching the target Nl . The intersection point of the rays with line i are $(x_i, z_i)_{i=1, \dots, Nl}$ and, $s_i = (-\sin \phi_i, \cos \phi_i)$ indicates the direction vector of the rays that leave i , with ϕ_i the angle that the ray forms with respect to the z -axis measured counterclockwise. As we consider only forward rays, the angles $\phi_i \in (-\pi/2, \pi/2)$. Therefore, a ray segment between (x_i, z_i) and (x_j, z_j) with $j \neq i$ is parameterized in real space by:

$$\mathbf{r}(s) = \begin{pmatrix} x_i - s \sin(\phi_i) \\ z_i + s \cos(\phi_i) \end{pmatrix} \quad 0 \leq s \leq s_{\max}, \quad (3.1.1)$$

where s denotes the arc-length and s_{\max} is the maximum value that it can assume. To make clear the notation, in Fig. we show an example where a single ray is traced inside a very simple optical system, the so-called two-faceted cup. The light source $S = [-a, a]$ (line 1) and the target $T = [-b, b]$ (line 4) are two segments normal to the z -axis, where $a = 2$ and $b = 17$. The left and right reflectors (line 2 and 3) are oblique

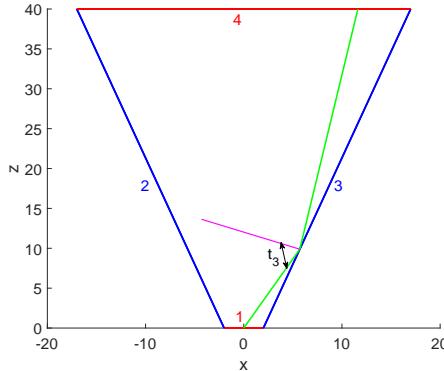


Figure 3.1: Shape of the two-faceted cup. Each line of the system is labeled with a number. The source $S = [-2, 2]$ (line number 1) is located on the x -axis. The target $T = [-17, 17]$ (line 4) is parallel to the source and is located at a height $z = 40$. The left and right reflectors (line 2 and 3) connect the source with the target.

segments that connect the source and the target. All the optical lines $i \in \{1, \dots, 4\}$ are located in air, therefore the refractive index $n_i = 1$ for every i .

3.2 Monte Carlo ray tracing

Assuming a Lambertian source, the input intensity at S emitted in the direction t_1 is given by:

$$I(t_1) = 2aL \cos(t_1), \quad (3.2.1)$$

where L is the luminance and a is the half width of S . In order to compute the target intensity, we need to find a relation between the intensities at S and T . Hence, we need to know how the optical system influences the direction of the rays when they hit an optical line. To this purpose, the ray tracing procedure is often used in optics. Ray tracing relates the position coordinates (x_1, z_1) and the direction vector s_1 of every ray at the source S with the corresponding position (x_4, z_4) and direction s_4 at the target T . As in the following we will use often the target coordinates of the rays, from now on, to simplify the notation, we write t instead of t_4 and (x, z) instead of (x_4, z_4) for the target coordinates.

The ray tracing algorithm can be schematized as follows. For every ray that leaves S with initial position (x_1, z_1) and initial angle t_1 , its ray parametrization is implemented according to Eq. (3.2.13). Then, the coordinates (x_i, z_i) of the intersection point between the ray and the line i that it hits are computed. The unit normal ν_i to the line i at the point (x_i, z_i) is calculated to compute the change of direction of the ray. Since all the lines of the system are located in air, only the reflection law plays a role, [?]. Therefore, denoting with t_1 the direction of the incident ray, the direction t_2 of the reflected ray is given by:

$$t_2 = t_1 - 2(t_1, \nu_i)\nu_i, \quad (3.2.2)$$

where the vectors \mathbf{t}_1 and \mathbf{t}_2 are unit vectors, [?]. The procedure explained above is repeated for every line that the ray encounters until it reaches the target and for every ray traced through the system.

There are different ways to implement the ray tracing procedure. An often used method is MC ray tracing which calculates the target photometric variables considering a sample of many rays that are traced randomly from S to T . The output intensity is computed as a function of the angular coordinate t and is calculated dividing the target into intervals of the same length, the so-called bins. A partitioning $P_1 : -\pi/2 = t_0 < t_1 < \dots < t_{Nb} = \pi/2$ of the interval $[-\pi/2, \pi/2]$ is defined where Nb is the number of bins in P_1 . We remark that, with a slight abuse of notation, we indicated the angular coordinates of the rays at the target with t_j instead of $t_{4,j}$ for every $j \in \{0, \dots, Nb\}$. The normalized approximated intensity $g_{MC}(t)$ is a piecewise constant function and its value over the j -th bin is the ratio between the number of rays that fall into that bin $Nr[t_{j-1}, t_j]$ and the total number of rays traced $Nr[-\pi/2, \pi/2]$. Hence, g_{MC} is defined by:

$$g_{MC}(t) = \frac{Nr[t_{j-1}, t_j]}{Nr[-\pi/2, \pi/2]} \quad \text{for } t \in [t_{j-1}, t_j]. \quad (3.2.3)$$

Furthermore, the output intensity is computed from the value of the intensity $g_{MC}(t_{j-1/2})$ along the direction $t_{j-1/2} = (t_{j-1} + t_j)/2$ for every bin $[t_{j-1}, t_j]_{j=1, \dots, Nb}$. The intensity $g_{MC}(t_{j-1/2})$ gives an estimate of the probability that a ray reaches the target with an angle in the j -th interval $[t_{j-1}, t_j]$ of the partitioning P_1 . This probability $P_{j,\Delta t}$ is given by:

$$P_{j,\Delta t} = \Pr(t_{j-1} \leq t < t_j) = \frac{\int_{t_{j-1}}^{t_j} G(t) dt}{\int_{-\pi/2}^{\pi/2} G(t) dt}, \quad (3.2.4)$$

where $G(t)$ is the output intensity (not normalized) and it is measured in lumen per radian [lm/rad]. Note that $\sum_{j=1}^{Nb} P_{j,\Delta t} = 1$. Using the mean value theorem for the function $G(t)$ continuous in $[t_{j-1}, t_j]$, the integral at the numerator of the previous equation can be written as:

$$\int_{t_{j-1}}^{t_j} G(t) dt = \Delta t G(t_{j-1/2}). \quad (3.2.5)$$

Hence, $P_{j,\Delta t}$ is proportional to the size $\Delta t = (t_{Nb} - t_0)/Nb$ of the intervals, i.e., inversely proportional to the number of bins Nb of the partitioning P_1 . Indicating with $\Phi = \int_{-\pi/2}^{\pi/2} G(t) dt$ the total flux (measured in lumen [lm]), the error between the intensity $G(t_{j-1/2})$ and the averaged MC intensity $\Phi g_{MC}(t_{j-1/2})/\Delta t$ is given by:

$$\begin{aligned} & \left| G(t_{j-1/2}) - \frac{\Phi}{\Delta t} g_{MC}(t_{j-1/2}) \right| \leq \\ & \left| G(t_{j-1/2}) - \frac{1}{\Delta t} \int_{t_{j-1}}^{t_j} G(t) dt \right| + \\ & \frac{1}{\Delta t} \left| \int_{t_{j-1}}^{t_j} G(t) dt - \Phi g_{MC}(t_{j-1/2}) \right|. \end{aligned} \quad (3.2.6)$$

The first term of the right hand side of inequality (3.2.6) gives an estimate of how much the averaged intensity $\frac{1}{\Delta t} \int_{t_{j-1}}^{t_j} G(t)dt$ differs from the exact intensity $G(t_{j-1/2})$. This term is due to the discretization of the target and therefore it depends on the number of bins Nb considered. Substituting $G(t)$ with its Taylor expansion around the point $t_{j-1/2}$ we obtain that this term is proportional to the square of the size of the bins, therefore the following equality holds:

$$\left| G(t_{j-1/2}) - \frac{1}{\Delta t} \int_{t_{j-1}}^{t_j} G(t)dt \right| = C_1/Nb^2 \quad (3.2.7)$$

with $C_1 > 0$ a certain constant.

The second part of the right hand side of inequality (3.2.6) gives an estimate of the MC error and therefore it depends also on the number of rays traced. In order to show how this term decreases as a function of the number of rays traced, we define the random variable $X_j(t)$ as the variable that is equal to 1 if the ray with angular coordinate t is inside the interval $[t_{j-1}, t_j]$ and equal to 0 otherwise,

$$X_j(t) = \begin{cases} 1 & \text{if } t \in [t_{j-1}, t_j], \\ 0 & \text{otherwise.} \end{cases} \quad (3.2.8)$$

The Bernoulli trial X_j follows a binomial distribution $B(1, P_{j,\Delta t})$. Considering a sample of Nr rays, the variable $Y_j = \sum_{k=1}^{Nr} X_j(t_k)$ follows a binomial distribution $B(Nr, P_{j,\Delta t})$, where t_k is the angle that the k -th ray forms with the optical axis. Then, using the de Moivre-Laplace theorem, we conclude that the variable Y_j is approximated by a normal distribution with mean value $E[Y_j] = NrP_{j,\Delta t}$ and variance $\sigma^2[Y_j] = NrP_{j,\Delta t}(1 - P_{j,\Delta t})$ when a large number of rays is considered, see [?, ?]. Thus, the normalized intensity along the direction $t_{j-1/2}$ is given by:

$$g_{MC}(t_{j-1/2}) = \sum_{k=1}^{Nr} X_j(t_k)/Nr. \quad (3.2.9)$$

The mean value $E[g_{MC}(t_{j-1/2})] = P_{j,\Delta t}$ and the variance $\sigma^2[g_{MC}(t_{j-1/2})] = P_{j,\Delta t}(1 - P_{j,\Delta t})/Nr$. Note that the standard deviation $\sigma_j := \sigma[g_{MC}(t_{j-1/2})]$ equals:

$$\sigma_j = \sqrt{P_{j,\Delta t}(1 - P_{j,\Delta t})/Nr} = \frac{C_2}{\sqrt{NbNr}}, \quad (3.2.10)$$

for some $C_2 > 0$. σ_j can be used to give an estimate of the difference between the intensity $g_{MC}(t_{j-1/2})$ and its mean value $P_{j,\Delta t}$. Therefore, the second term of the right hand side of relation (3.2.6) becomes:

$$\begin{aligned} \frac{1}{\Delta t} \left| \int_{t_{j-1}}^{t_j} G(t)dt - \Phi g_{MC}(t_{j-1/2}) \right| &= \\ \frac{\Phi}{\Delta t} \left| P_{j,\Delta t} - g_{MC}(t_{j-1/2}) \right| &\propto \\ \frac{\Phi}{\Delta t} \sigma_j [g_{MC}(t_{j-1/2})] &= C_3 \frac{Nb}{\sqrt{NbNr}} = C_3 \sqrt{\frac{Nb}{Nr}}, \end{aligned} \quad (3.2.11)$$

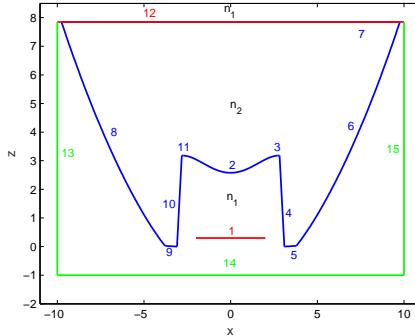


Figure 3.2: Shape of the TIR-collimator. Each line of the system is labeled with a number. The source $S = [-2, 2]$ (line number 1) is located at an height $z_s = 0.3$ from the x -axis. The target $T = [-9.7, 9.7]$ (line 12) is parallel to the source and is located at an height $z_t = 7.85$. The shape of the collimator is shown as a blue line. Three detectors depicted with green lines (lines 13, 14, and 15) are located at the left, the right and the bottom of the optical system. $n_1 = 1$ is the refraction index of the medium (air) where the source and the target are located, and $n_2 = 1.5$ the refraction index of the medium (glass) inside the optical system. The sagitta of the lens is equal to 0.6.

for some $C_3 > 0$, where the approximation holds because σ_j gives a measure for the error between $g_{\text{MC}}(t_{j-1/2})$ and the probability $P_{j,\Delta t}$, [?]. The second equality follows from Eq. (3.2.10). To conclude, the MC error over the j -th bin is estimated by:

$$\left| G(t_{j-1/2}) - \frac{\Phi}{\Delta t} g_{\text{MC}}(t_{j-1/2}) \right| = \frac{C_1}{Nb^2} + C_4 \sqrt{\frac{Nb}{Nr}}, \quad (3.2.12)$$

for $C_4 > 0$. Considering a fixed number of rays, we obtain that the minimal error is reached when $Nb \approx Nr^{1/5}$. Hence, if 10^{10} rays are considered the target has to be divided into 10^2 bins to minimize the MC error. This leads to computational efforts resulting in a very slow procedure.

To obtain the photometric variables at the target, the propagation of light from the source S to the target T is computed. In this work, we calculate the output intensity for the TIR-collimator, the profile of which is depicted in Figure 3.2. The TIR-collimator we analyze is an optical system rotationally symmetric with respect to the z -axis and consists of a lens (line 2), two broken lines adjacent to the lens (formed by the collection of the segments 3, 4, and 5 and 9, 10 and 11), two curved lines (labeled with 6 and 8) and the top formed by a horizontal segment (line 7). The lens and the broken lines are refractive lines while, the curved lines are designed in such a way that light is internally reflected (which explains the name TIR). The light source S (line 1) and the target T (line 12) are two straight segments normal to the optical axis and are located in air ($n_1 = 1$) while the volume inside the collimator is filled with a material with index of refraction $n_2 = 1.5$ (e.g. glass). The collimator is surrounded by two vertical and two horizontal lines (lines 13, 15, 12 and 14, respectively) that receive the light exiting from the optical system; among these, the horizontal one at the top is assumed to be the light target and, it is located at a small distance from the top.

From now on, the coordinates $(x_i, z_i)_{(i=1, \dots, 15)}$ denote the intersection of the rays with the line i and, $\mathbf{s}_i = (-\sin t_i, \cos t_i)$ is the direction vector of the rays that leave the line i , where $t_i \in (-\pi/2, \pi/2)$ is the angle that the ray forms with the z -axis, measured counterclockwise. Therefore, a ray segment between (x_i, z_i) and (x_{i+1}, z_{i+1}) is parameterized by:

$$\mathbf{r}(s) = \begin{pmatrix} x_i - s \sin(t_i) \\ z_i + s \cos(t_i) \end{pmatrix} \quad s \geq 0, \quad (3.2.13)$$

where s denotes the arc-length.

A Lambertian optical source is considered; hence, the intensity over an interval $J = [-a, a]$ emitted in the direction t is given by:

$$I(t) = I_0 \cos(t), \quad (3.2.14)$$

where $I_0 = 2aL$, L is the luminance, and t is the angle that the ray forms with respect to the optical axis, measured counterclockwise. As a result, in the case where $L = 1$, I_0 coincides with the source length; from Equation (3.2.14), the intensity at the source is deduced.

To compute the target intensity, we need to know how the optical system changes the direction of the rays during their propagation from the source to the target. To do this, we employ the ray tracing technique which can be summarized as follows: first, a ray from the source with initial position given by the coordinates (x_1, z_1) and initial angle t_1 with respect to the z -axis is traced and, the ray parametrization is implemented according to Equation (3.2.13). Second, the intersection point (x_i, z_i) between the ray and the line i that it hits first is computed. Third, the normal to the line hit at the point (x_i, z_i) is calculated to compute the change of direction of the ray. For the last step, the laws of reflection and refraction are implemented. The direction of the refractive ray is given by:

$$\mathbf{t} = n_{1,2} \mathbf{i} + \left[\sqrt{1 - n_{1,2}^2 + n_{1,2}^2(\mathbf{n}, \mathbf{i})^2} - n_{1,2}(\mathbf{n}, \mathbf{i}) \right] \mathbf{n}, \quad (3.2.15)$$

where $n_{1,2} = n_1/n_2$ with n_1 and n_2 the refraction indexes of air and of glass, respectively. The unit vectors \mathbf{i} and \mathbf{t} describe the directions of the incident and refracted ray, respectively; \mathbf{n} is the normal to the line; it is also a unit vector and it is directed towards the interior of the optical system. Note that the positive sign before the square root is due to the convention to take the inward direction of the normal \mathbf{n} (see [8], chapter 4, p. 95-106, and, [?], chapter 12 p. 403-409). In the case where $n_1 = -n_2$, Equation (3.2.15) can be rewritten as the law of reflection:

$$\mathbf{t} = \mathbf{i} - 2(\mathbf{i}, \mathbf{n})\mathbf{n}. \quad (3.2.16)$$

Equation (3.2.16) is used when the total internal reflection condition holds, that is when the following inequality is true:

$$1 - n_{1,2}^2 + (\mathbf{n}, \mathbf{i})^2 < 0. \quad (3.2.17)$$

For the TIR-collimator the previous condition occurs for the curved lines (lines 6 and 8 in Figure 3.2). Finally, the new parametrization of the ray is described by:

$$\mathbf{r}(s) = \begin{pmatrix} x_i + s t_x \\ z_i + s t_z \end{pmatrix}, \quad (3.2.18)$$

where t_x and t_z are the x and z -components of the new ray direction and are calculated from Equations (3.2.15) or (3.2.16). The points (x_i, z_i) and the new direction \mathbf{t} are computed until the ray hits the target and the previous procedure is repeated for each ray traced.

To obtain a reasonable approximation of the target intensity, a large number of rays has to be traced; the more rays are traced, the more accurate the target intensity is. Moreover, for the TIR-collimator shown in Figure 3.2, we do not have an explicit equation to describe the reflectors. Only the positions of a discrete set of points located on their curves are known. Therefore, we use spline interpolation to obtain a good approximation of the curved lines. In addition, to calculate the intersection points between the rays and these lines, the Newton-Raphson procedure is employed. Due to all these reasons, the ray-tracing method is a very slow procedure.

A frequently used ray tracing method in non-imaging optics is MC ray tracing [11] in which the rays are emitted from a random location and at random angle. They are traced through the system until they reach the target receiver. To calculate the output intensity, the target screen is divided into bins and the frequency of the rays that arrive at each bin is considered. The intensity restricted to a certain bin is obtained by dividing the number of rays that fall into that bin by the total number of rays traced. Although MC ray tracing is highly robust and does not require difficult calculations, it has two main disadvantages. First, some information is lost because the flux of a ray is averaged over a bin. Second, some parts of the target are reached by a very small fraction of rays and, consequently, the intensity is unreliable in those parts. As a consequence, a large number of rays needs to be traced to obtain an accurate intensity making the MC method computationally expensive.

We provide a new method that employs the phase space representation of the optical system to avoid tracing rays where the luminance does not present any discontinuities. Phase space ray tracing is explained in the next section.

In phase space each ray is described by its intersection point with the line it hits and the sine of the angle it forms with respect to the optical axis multiplied by the refractive index (see [12] chapter 2.1-2.3, [?], and [?] chapter 1 for details). In the following, the phase space is considered only for the source \mathcal{S} and the target \mathcal{T} and for no other line of the optical system. The rays in a two-dimensional system correspond to points with coordinates (x, τ) and (q, η) in \mathcal{S} and \mathcal{T} phase space, respectively. We have indicated the ray positions with x and q , the angles formed with the normal with t and θ , the refractive indexes with n_s and n_t , for \mathcal{S} and \mathcal{T} , respectively and, with $\tau = n_s \sin(t)$ and $\eta = n_t \sin(\theta)$ the directions of the rays.

The rays are represented by a unique point in phase space, both for \mathcal{S} and \mathcal{T} . More formally, the optical phase space for the light source is defined as:

$$\mathcal{P}_s = \mathcal{S} \times [-n_s, n_s]. \quad (3.2.19)$$

The target phase space is defined as

$$\mathcal{P}_t = \mathcal{T} \times [-n_t, n_t]. \quad (3.2.20)$$

The map $\mathcal{M} : \mathcal{P}_s \rightarrow \mathcal{P}_t$ which describes how the optical system changes the rays is defined as:

$$\mathcal{M}(x, \tau) = (q, \eta). \quad (3.2.21)$$

For most optical systems, there is no way to determine an explicit expression for the map \mathcal{M} defined above. The idea is to apply the edge-ray principle [13] to a given set of rays at the source. The principle states that to map one region from the source to the target phase space it is sufficient to map the boundaries of those regions. Therefore, the boundaries of the source are mapped to the boundaries of the target and the regions where the luminance is different from zero are calculated. The intensity in target phase space is defined as a function of the output luminance:

$$I_{PS}(\eta) = \int_{\mathcal{T}_\eta} L_t(q, \eta) dq, \quad (3.2.22)$$

where, for a given constant $\eta_0 \in [-1, 1]$, the set $\mathcal{T}_{\eta_0} = \{(q, \eta) \in \mathcal{T} \mid \eta = \eta_0\}$ and $L_t(q, \eta)$ indicates the luminance at the target. As we use the target phase space to compute the output intensity, it is convenient to define it as a function of $\sin(\theta)$ instead of θ . Note that the luminance is positive in the entire \mathcal{P}_s , but not all parts of \mathcal{P}_t receive light emitted by the source. As a result, L_t has jump discontinuities where it changes from zero to positive values. To understand where these discontinuities occur, further information about the rays is required. Because of this, for PS ray tracing not only the initial positions and the initial angles of the rays are stored, but also the optical lines they hit when they propagate through the system. A ray path Π is defined as the collection of lines hit by the ray. Rays that are close to each other at the source and leave the source at close angles follow the same path and hit the target at close positions and under close angles. All the rays that follow the same path are grouped together into the same subset of phase space. From now on, we indicate with p the number of all the possible paths $(\Pi_j)_{j=1, \dots, p}$ encountered by the rays and, with R_{s, Π_j} and R_{t, Π_j} the regions corresponding to the rays that follow the path Π_j for the source and the target, respectively. The map \mathcal{M} defined in Equation (3.2.21) relates the regions R_{s, Π_j} to the regions R_{t, Π_j} for every $j \in \{1, \dots, p\}$. The edge-ray principle guarantees that the boundaries $\partial R_{s, \Pi_j}$ and $\partial R_{t, \Pi_j}$ are connected by the same map \mathcal{M} . Given two different paths Π_1 and Π_2 , the regions R_{t, Π_1} and R_{t, Π_2} do not overlap; they can have at most a common boundary. As a result, the discontinuities of the luminance occur exactly at the boundaries $(\partial R_{t, \Pi_j})_{j=1, \dots, p}$. Finally, the luminance at the target satisfies the following relations:

$$\begin{aligned} L_t(q, \eta) &> 0 && \text{for } (q, \eta) \in (R_{t, \Pi_j})_{j=1, \dots, p}, \\ L_t(q, \eta) &= 0 && \text{otherwise.} \end{aligned} \quad (3.2.23)$$

In addition, the luminance is conserved along a ray, so it remains constant inside every region $(R_{t, \Pi_j})_{j=1, \dots, p}$, (see [?], chapter 16). The output intensity is obtained from Equation (3.2.22). Therefore, the problem to compute the target intensity can be interpreted as the calculation of the boundaries $(\partial R_{t, \Pi_j})_{j=1, \dots, p}$. To this end, we define a triangulation on source phase space in such a way that more rays close to the boundaries are traced. The details of these procedures are explained in the next section.

3.2.1 Triangulation refinement of source phase space

The regions $(R_{t, \Pi_j})_{j=1, \dots, p}$ can be defined only when some rays are traced. Given an initial set of rays, the rays closest to the boundaries $(\partial R_{t, \Pi_j})_{j=1, \dots, p}$ are selected

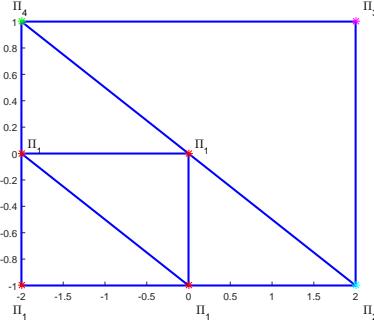


Figure 3.3: Triangulation refinement: when the rays related to the vertices of the triangles follow a different path a new refinement step is required. Each refinement step leads to four new triangles. The parameters values are $\epsilon_{x_{max}} = 2$, $\epsilon_{\tau_{max}} = 1$, $\epsilon_{x_{min}} = 4$ and $\epsilon_{\tau_{min}} = 2$.

and more rays in their vicinity are created to get progressively better estimates of the boundaries. A more detailed description is provided below. A triangulation in \mathcal{P}_s is defined and a ray from every vertex (x_k, τ_k) of the triangle is traced. The procedure starts tracing four rays with coordinates $(x_k, \tau_k)_{k=1, \dots, 4}$ that are located exactly at the corners of \mathcal{P}_s and, for each of them, the paths $(\Pi_j)_{j=1, \dots, 4}$, are stored. Next, for some $j \in \{1, \dots, 4\}$, the grid is divided into two equal triangles joining two opposite vertices. For each triangle the rays located at its corners are traced. If the paths corresponding to those rays are different, one or more boundaries $(\partial R_{t, \Pi_j})_{j=1, \dots, 4}$ are expected to cross the triangle. In that case, the middle points $(x_k, \tau_k)_{k=5, 6, 7}$ of each side of the triangle are added and three more rays with coordinates $(x_k, \tau_k)_{k=5, 6, 7}$ are traced. Each refinement step leads to four new triangles (see Figure 3.3).

When all the rays in the corners of each triangle have the same path, it is not necessary to refine the triangles anymore. Note that it can happen that a region formed by rays that follow a path Π_j is located completely inside a triangle whose vertices are related to the same path Π_i with $j \neq i$. In that case the algorithm is not able to detect that region, see Figure 3.4. To avoid this, two parameters $\epsilon_{x_{min}}$ and $\epsilon_{\tau_{min}}$ are defined for the x -axis and the τ -axis, respectively. When the length of the sides of the triangle are greater than these parameters, a new triangle is defined even if its vertices correspond to the same path. Furthermore, two other parameters $\epsilon_{x_{max}}$ and $\epsilon_{\tau_{max}}$ are introduced to define a stopping criterion. The algorithm stops when the length of the sides of the triangles is smaller than $\epsilon_{x_{max}}$ and $\epsilon_{\tau_{max}}$. The values of the parameters $\epsilon_{x_{max}}$, $\epsilon_{\tau_{max}}$, $\epsilon_{x_{min}}$ and $\epsilon_{\tau_{min}}$ determine the number of rays traced. Indeed, on the one hand, $\epsilon_{x_{max}}$ and $\epsilon_{\tau_{max}}$ can be decreased to obtain more rays close to the boundaries; on the other hand, a large number of rays in the interior of the regions can be traced decreasing the values of $\epsilon_{x_{min}}$ and $\epsilon_{\tau_{min}}$.

Using the above procedure, rays increasingly closer to the boundaries are traced. For our optical system, the width of the x -axis in source phase space is two times the width of the τ -axis. Thus, our choice is $\epsilon_{\tau_{min}} = \frac{1}{2}\epsilon_{x_{min}}$ and $\epsilon_{\tau_{max}} = \frac{1}{2}\epsilon_{x_{max}}$. Figure 3.5 shows an example of a triangulation refinement of the source phase space with $\epsilon_{x_{max}} = 0.1$ and $\epsilon_{x_{min}} = 1$. The triangulation refinement provides more triangles close to the boundaries $\partial R_{s, \Pi_j}$ than those inside the regions R_{s, Π_j} .

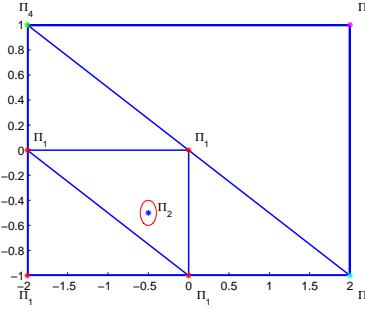


Figure 3.4: The red line encloses a region of rays that follow the path Π_2 and is completely located inside a triangle. The algorithm is not able to detect that region and, a further refinement is required. The parameters values are $\epsilon_{x_{max}} = 2$, $\epsilon_{\tau_{max}} = 1$, $\epsilon_{x_{min}} = 4$ and $\epsilon_{\tau_{min}} = 2$.

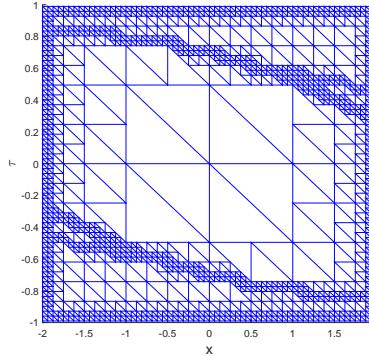


Figure 3.5: Triangulation refinement of source phase space: near the boundaries more rays are traced. The values of the parameters are $\epsilon_{x_{max}} = 0.1$ and $\epsilon_{x_{min}} = 1$.

The paths $(\Pi_j)_{j=1,\dots,p}$ followed by the rays located at the corner of the triangles are computed during the procedure and, the regions R_{s,Π_j} and R_{t,Π_j} are defined for each Π_j . Next, a criterion to select the values of the parameters $\epsilon_{x_{min}}$ and $\epsilon_{x_{max}}$ and a method to compute the boundaries $\partial R_{t,\Pi_j}$ is provided. Furthermore, the output photometric variables are computed, the details are explained in the next section. As mentioned in Section ???, the boundaries $(\partial R_{t,\Pi_j})_{j=1,\dots,p}$ have to be calculated to compute the photometric variables at the target. Our method is based on the triangulation refinement of the source phase space. More rays close to the boundaries can be traced selecting increasingly smaller values for the parameters $\epsilon_{x_{max}}$ and $\epsilon_{\tau_{max}}$. Once the algorithm stops, only the triangles that are expected to be crossed by a boundary are taken into account. By construction, each of these triangles has two vertices that follow the same path and one vertex that follows another path. The triangles are ordered in such a way that two of them are neighbors if they have a side in common. Given a path Π_j with $j \in \{1, \dots, p\}$ the boundary $\partial R_{s,\Pi_j}$ of the region corresponding to Π_j is approximated by those vertices of the triangles corresponding

to the path Π_j . The boundaries $\partial R_{t,\Pi_j}$ at the target are given by $\mathcal{M}(\partial R_{s,\Pi_j})$ for every $j \in \{1, \dots, p\}$. To establish the minimum value of the parameter $\epsilon_{x_{max}}$ that gives a good approximation of the boundaries $\partial R_{t,\Pi_j}$, a technique that exploits the conservation of the étendue in phase space is provided, (see [?], chapter 16). The essence of our approach is as follows.

We consider the étendue for the whole \mathcal{P}_s which is given by:

$$E_s = 2n_s a \sin(t_{max}), \quad (3.2.24)$$

where a is the length of the source and t_{max} is the maximum value of the angle that the rays make with the z -axis. The étendue of a set of rays is defined by the area they occupy in phase space. For the TIR-collimator we considered (Figure 3.2), the area of \mathcal{P}_s is equal to 7.92, as $a = 4$ and $\sin(t_{max}) = 0.99$. The rays traced are uniformly distributed over \mathcal{P}_s , and they cover it entirely.

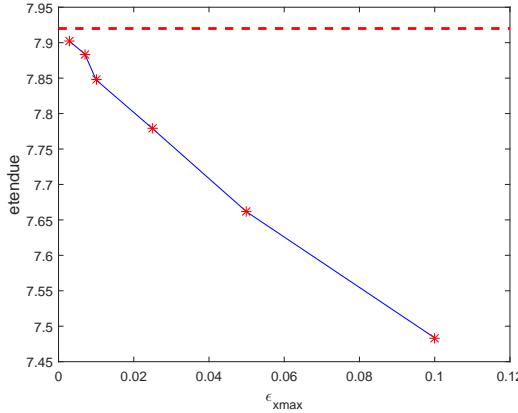


Figure 3.6: The total étendue as an area in PS is depicted with the dotted red line. The approximated étendue is computed for a range of values of $\epsilon_{x_{max}}$. Decreasing the value of the parameter the $\epsilon_{x_{max}}$, the étendue increases and reaches a good approximation of the exact étendue for $\epsilon_{x_{max}} = 3 \cdot 10^{-3}$.

The total étendue at the target E_t is given by the sum of the étendues related to each region R_{t,Π_j} :

$$E_t = \sum_{j=1}^p E(R_{t,\Pi_j}), \quad (3.2.25)$$

where $E(R_{t,\Pi_j})$ is the contribution to the étendue at the target given by the rays inside the region R_{t,Π_j} . Note that E_t is computed by also considering the area of the regions formed by the rays that hit the left and the right detectors (lines 13 and 14 in Figure 3.2). $E(R_{t,\Pi_j})$ is defined by:

$$E(R_{t,\Pi_j}) = \iint_{R_{t,\Pi_j}} dq d\eta. \quad (3.2.26)$$

To calculate the previous integral the triangulation refinement method is applied to the regions R_{s,Π_j} for a range of values of $\epsilon_{x_{max}}$, with an approximation of the boundaries $\partial R_{s,\Pi_j}$ obtained for each of them. Therefore, the boundaries $\partial R_{t,\Pi_j}$ are also computed and the intersection points $(q_{\Pi_j,i}(\eta))_{i=1,\dots,r}$ between $\partial R_{t,\Pi_j}$ and the horizontal line $\eta = \text{const}$ are calculated for each $j \in \{1, \dots, p\}$, with $\eta \in [-1, 1]$. Ordering the points $(q_{\Pi_j,i}(\eta))_{i=1,\dots,r}$ in ascending order, Equation (3.2.26) becomes:

$$E(R_{t,\Pi_j}) = \sum_{i=1}^m \int_{-1}^1 (q_{\Pi_j,2i}(\eta) - q_{\Pi_j,2i-1}(\eta)) d\eta, \quad (3.2.27)$$

where m is the integer part of $r/2$ and r is the number of the intersection points between $\partial R_{t,\Pi_j}$ and the horizontal lines $\eta = \text{const}$. The integral in Equation (3.2.27) is calculated by discretizing the interval $[-1, 1]$ into $Nb = 100$ sub-intervals of equal width and using the trapezoidal rule. Figure 3.6 shows that decreasing the value of the parameter $\epsilon_{x_{max}}$ increases the values of the étendue at the target E_t , which reaches 7.9 when $\epsilon_{x_{max}} = 0.3 \cdot 10^{-3}$. We decide to stop the phase space refinement procedure when a good approximation of the étendue is obtained. Moreover, a criterion to establish the values of $\epsilon_{x_{min}}$ is provided. For each value of $\epsilon_{x_{max}}$ the étendue for a range of values of $\epsilon_{x_{min}}$ is computed. As the computation of the boundaries does not depend on the number of rays inside the regions, the étendue remains constant when the value of $\epsilon_{x_{min}}$ changes. We choose $\epsilon_{x_{min}}$ as large as possible avoiding to trace rays that do not significantly contribute to the computation of the photometric variables at the target. The value of $\epsilon_{x_{min}}$ depends on the distribution of the rays in phase space. For our optical system the parameter $\epsilon_{x_{min}} = 1$. Figure 3.7 and 3.8 show the approximation of the boundaries obtained for a set of $6.9 \cdot 10^4$ rays. Five different paths are found and rays that follow the same path are depicted with the same color. Figure 3.7 shows that, choosing the values of the parameters as explained above, the regions R_{s,Π_j} almost completely cover the source phase space. As a consequence, the dark areas in Figure 3.8 correspond to parts of target phase space that are not reached by any ray that leaves the source and propagates through the meridional plane of the optical system. Note that, using the triangulation procedure explained in the previous section, more rays close to the boundaries are traced.

To conclude, we compute the target intensity which is defined in \mathcal{P}_t by Equation (3.2.22). From Equation (3.2.23), we obtain:

$$I_{PS}(\eta) = \sum_{i,j} \int_{q_{\Pi_j,2i-1}(\eta)}^{q_{\Pi_j,2i}(\eta)} L_t(q, \eta) dq, \quad (3.2.28)$$

where the summation for the indices i is over all $i = 1, 2, \dots, m$, and the summation for the j indices is over all j for which the intersection between $\eta = \text{const}$ and R_{t,Π_j} is not empty. The intensity is expressed as a function of the angular parameter $\eta = n_t \sin(\theta)$. In the case of a Lambertian source with $L_s(x, \tau) = 1$, the following relation for the intensity at the target holds:

$$I_{PS}(\eta) = \sum_{i,j} (q_{\Pi_j,2i}(\eta) - q_{\Pi_j,2i-1}(\eta)), \quad (3.2.29)$$

where the relation $n_t = 1$ and conservation of luminance along a ray are exploited, (see [?], chapter 16). We again notice that Equation (3.2.28) and (3.2.29) are valid when

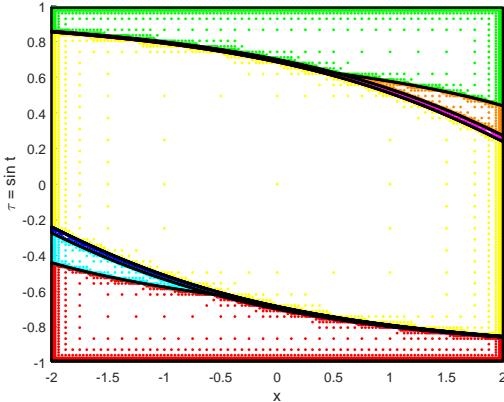


Figure 3.7: Distribution of the rays on source phase space. Around $6.9 \cdot 10^4$ rays are traced using the triangulation refinement with parameters: $\epsilon_{x_{min}} = 1$, $\epsilon_{\tau_{min}} = 0.5$, $\epsilon_{x_{max}} = 3 \cdot 10^{-3}$, $\epsilon_{\tau_{max}} = 1.5 \cdot 10^{-3}$. Rays that belong to the same region are depicted with the same color. The yellow rays follow the path $\Pi_1 = (1, 2, 7, 12)$; the red rays follow the path $\Pi_2 = (1, 10, 8, 7, 12)$; the green rays follow the path $\Pi_3 = (1, 4, 6, 7, 12)$; the blue rays follow the path $\Pi_4 = (1, 11, 7, 12)$; the magenta rays follow the path $\Pi_5 = (1, 3, 7, 12)$, the cyan rays hit the left detector (line 13) and follow the path $\Pi_6 = (1, 10, 7, 8, 13)$ and, the orange rays hit the right detector (line 15) and follow the path $\Pi_7 = (1, 4, 7, 6, 15)$. Each number corresponds to a line of the TIR-collimator as shown in Figure 3.2. The boundaries are depicted with the black lines.

only two intersection points are found. If $r > 2$ intersection points occur the sum of the distances $(q_{2i} - q_{2i-1})_{i=1, \dots, m}$ needs to be computed. To calculate the intensity for all the possible directions, a uniform partitioning $P : -1 \leq \eta_0 < \eta_1 \cdots < \eta_{Nb} \leq 1$ of the interval $J = [-1, 1]$ is considered, where $Nb = 100$. Eventually, the intensity for each η_h , with $h = 0, 1, \dots, Nb$, is obtained using relation (3.2.29). We compare the new method with the already existing MC ray tracing to show its efficiency.

The intensity for MC ray tracing is computed as follows. The partitioning P of $J = [-1, 1]$, used for the target phase space, is considered and the number of rays that fall into each bin $([\eta_h, \eta_{h+1}])_{h=0, \dots, Nb-1}$ is calculated for all $h \in \{0, \dots, Nb-1\}$. The intensity in the direction $\eta_k \in [\eta_h, \eta_{h+1}]$ is approximated by:

$$\hat{I}_{MC}(\eta_k) = \frac{Nr([\eta_h, \eta_{h+1}])}{Nr([-1, 1])}, \quad (3.2.30)$$

for every $\left(\eta_k = \frac{1}{2}(\eta_{h+1} + \eta_h)\right)_{k=1,2,\dots,Nb}$, where we have indicated the number of rays that fall into the bin $[\eta_h, \eta_{h+1}]$ with $Nr([\eta_h, \eta_{h+1}])$ and the total number of rays with $Nr([-1, 1])$. As \hat{I}_{MC} is normalized, a normalization of I_{PS} is also required to compare the two intensities. This normalization is calculated dividing the intensity by the étendue $E_{\mathcal{T}}$ at the target:

$$\hat{I}_{PS}(\eta_k) = \frac{1}{E_{\mathcal{T}}} \int_{\eta_h}^{\eta_{h+1}} I_{PS}(\eta) d\eta \quad \text{for } k = 1, 2, \dots, Nb, \quad (3.2.31)$$

where $E_{\mathcal{T}}$ is obtained by removing the étendue corresponding to the regions formed by

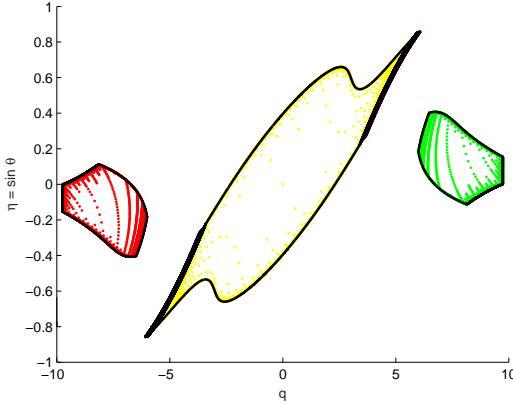


Figure 3.8: Target phase space representation of a set of $6.9 \cdot 10^4$ rays. Only the rays that hit the target (line 12) are considered. The values of the parameters $\epsilon_{x_{min}}, \epsilon_{x_{max}}, \epsilon_{\tau_{min}}, \epsilon_{\tau_{max}}$ and the choice of the colors for each path are the same as in Figure 3.7. The boundaries $(\partial R_{t,\Pi_j})_{j=1,\dots,p}$ are computed through the triangulation method. The dark areas correspond to areas that are not hit by any meridional plane.

the rays that hit the left and the right detectors from the total étendue E_t , computed in Equation (3.2.25) and shown in Figure 3.6. Note that the intensities are vectors of length Nb , and $\hat{I}_{PS}(\eta_k)_{k=1,2,\dots,Nb}$ represent the intensities along the directions $(\eta_k)_{k=1,\dots,Nb}$. The accuracy of the intensity also depends on the number of bins considered in the partitioning P . Choosing $Nb = 100$ results in a smooth profile of the intensity; hence, we decide to fix that value of Nb . The photometric variables at the target are now determined for MC and PS method. The numerical results are shown in the next section.

In this section a comparison between the MC and PS methods is presented. The MC and PS intensities are calculated several times increasing the number of rays Nr to improve the accuracy. Both approximate intensities are compared with an intensity taken as a reference. For some optical systems, there is an explicit solution for the target intensity but this is not the case for the TIR-collimator. Therefore, the reference intensity \hat{I}_{ref} is obtained considering $1,7 \cdot 10^8$ rays in the MC simulation. We show how the error, defined as:

$$\text{error} = \frac{\sum_{h=1}^{Nb} |\hat{I}_{PS}(\eta_h) - \hat{I}_{ref}(\eta_h)|}{Nb}, \quad (3.2.32)$$

decreases with the increase in the number of rays. Table 3.1 describes how the number of rays traced affects the error estimation and shows the correlation between étendue and the number of rays, which is determined by the values of $\epsilon_{x_{min}}, \epsilon_{\tau_{min}}, \epsilon_{x_{max}}$ and $\epsilon_{\tau_{max}}$ as explained in Section ??3.2.1. Next, the intensity \hat{I}_{MC} for the MC method is computed. Replacing \hat{I}_{PS} with \hat{I}_{MC} in Equation (3.2.32), the error between the reference intensity and the MC intensity is calculated. Increasing the number of rays traced in MC ray tracing, the error gradually decreases. In Table 3.2 the numerical results are reported.

Table 3.1: Error values of the PS intensity

Number of rays	$\epsilon_{x_{max}}$	$\epsilon_{\tau_{max}}$	étendue	error
1 403	$1.0 \cdot 10^{-1}$	$5.00 \cdot 10^{-2}$	7.4836	$3.57 \cdot 10^{-4}$
3 237	$5.0 \cdot 10^{-2}$	$2.50 \cdot 10^{-2}$	7.6614	$2.22 \cdot 10^{-4}$
7 299	$2.5 \cdot 10^{-2}$	$1.25 \cdot 10^{-2}$	7.7787	$1.38 \cdot 10^{-4}$
15 919	$1.0 \cdot 10^{-2}$	$5.00 \cdot 10^{-3}$	7.8475	$7.31 \cdot 10^{-5}$
33 651	$7.0 \cdot 10^{-3}$	$3.50 \cdot 10^{-3}$	7.8839	$3.80 \cdot 10^{-5}$
69 330	$3.0 \cdot 10^{-3}$	$1.50 \cdot 10^{-3}$	7.9017	$2.02 \cdot 10^{-5}$

Table 3.2: Error values of the MC intensity

Number of rays	error _{MC}
970	$2.20 \cdot 10^{-3}$
9 702	$6.60 \cdot 10^{-4}$
97 104	$1.74 \cdot 10^{-4}$
971 436	$6.34 \cdot 10^{-5}$
9 715 391	$2.06 \cdot 10^{-5}$

The results listed in Table 3.1 and Table 3.2 are shown in Figure 3.9, where the red line depicts the behavior of the error for the PS intensity, and the blue line indicates the error for the MC simulation.

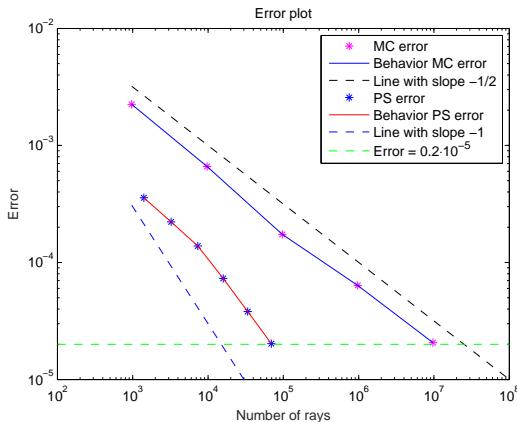


Figure 3.9: The red line depicts the error between the intensity on phase space and the reference intensity. The blue line shows the error between the Monte Carlo intensity and the reference intensity. The dashed black line represents a straight line with the slope equal to $-\frac{1}{2}$. The dashed blue line represents a straight line with the slope equal to -1 . The horizontal dotted line shows that an error equal to $2.00 \cdot 10^{-5}$ can be obtained tracing at least 10^2 times fewer rays in phase space.

Figure 3.9 shows that an error equal to $2.00 \cdot 10^{-5}$ is obtained by tracing around $9.7 \cdot 10^6$ rays for MC and only around $6.9 \cdot 10^4$ in PS. The error decreases as $\frac{1}{\sqrt{N_r}}$ for

the MC method and as $\frac{1}{Nr}$ for the PS simulation.

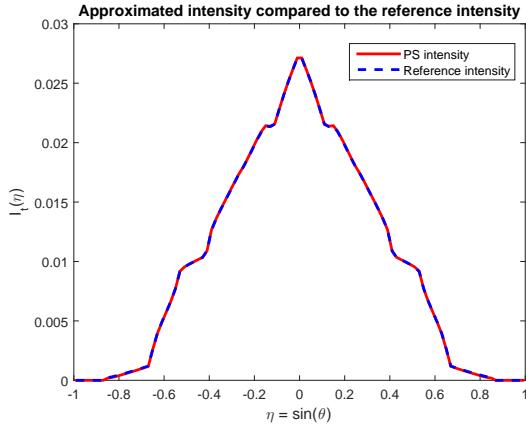


Figure 3.10: The red line shows the PS intensity at the target of the TIR-collimator. The reference intensity is depicted with the dotted blue line. The approximate intensity can hardly be distinguished from the exact intensity. The curves are functions of the angular parameter $\eta = n_t \sin(\theta)$.

The intensity profile $\hat{I}_{PS}(\eta)$ obtained with the phase space method and tracing around $6.9 \cdot 10^4$ rays is depicted in Fig. 3.10 with a red line. \hat{I}_{PS} is hardly distinguishable from \hat{I}_{ref} (dashed blue line in Figure 3.10). The intensities are expressed as functions of η and thus, the curves in Figure 3.10 do not depict the spatial intensities.

Finally, we claim that PS ray tracing is also more accurate than the ray tracing procedure proposed by Moore (2013), [?]. The novelty of our approach compared to the method used by Moore, is briefly explained below. First, to compute the output intensity, we employ the phase space of the target. This avoids the use of any interpolation to compute the photometric variables and therefore, more accurate results are obtained. Second, in [?] all rays that leave the source start at the same position and only a sampling angular range is given. In our approach a rectangular source is considered thus, both the angular and spatial coordinates of each ray change. This extra variable can produce very irregular shapes of the regions at target phase space. To overcome this issue, we employ the edge-ray principle and we consider the regions at source phase space where the distribution of the rays is much more regular and the corresponding boundaries are easily computed. As a consequence, our procedure is suitable to compute the output intensity as function of both the angular or the spatial coordinates. Third, using the conservation of étendue, we provided a criterion to stop the triangulation refinement. In this way we can estimate the number of rays required to obtain the desired accuracy and thus, we avoid tracing more rays than necessary.

3.3 Quasi-Monte Carlo method

Chapter 4

Ray tracing on phase space

4.1 Phase space concept

4.2 The edge-ray principle

4.3 Phase space ray tracing

Chapter 5

Two different approaches to compute the boundaries in target phase space

5.1 The α -shapes approach

Given a finite set \mathcal{S} of points we want to determine the shape formed by these points. α -shapes are geometrical objects which give us a good approximation of the shape of a given point set \mathcal{S} . Before giving a formal definition we explain an intuitive interpretation of α -shapes. As mentioned in [14] we can think of an α -shape as a mass of ice-cream with several chocolate pieces. The mass making up the space \mathbb{R}^3 and the chocolate pieces are the point set \mathcal{S} . Then the aim is to find the shape formed by the chocolate pieces. We can use a spoon with a spherical shape and carve out all parts of the ice-cream without removing the chocolate pieces. We will obtain a shape formed by arcs and points (see figure below for the two-dimensional case). Straightening the arcs to triangles and line segments we have an intuitive description of what is called the α -shape of \mathcal{S} . In our example, the parameter α determines of the radius of the carving spoon. If α is equal to 0 the shape degenerates to the point set \mathcal{S} . On the other hand, when $\alpha \rightarrow \infty$ the α -shape is simple the convex hull. More precisely the process is summarized as follows. Given a point cloud \mathcal{S} we start with a triangulation of it (a possible choice could be the Delaunay triangulation described in the next section). For each triangle we calculate the radius of the circumcircle. If the radius is larger than α the triangle is removed from the shape. The rule of the parameter α is highly significant in this procedure. Hence we have to choose it in such a way to get a better approximation. The choice of the parameter α is closely related to the radius of the circumcircles. A possible strategy is to find the radius of the greater empty circumcircle. Thus α is related to the density of the points. In particular we have:

$$\alpha = C \frac{1}{\Delta} , \quad (5.1.1)$$

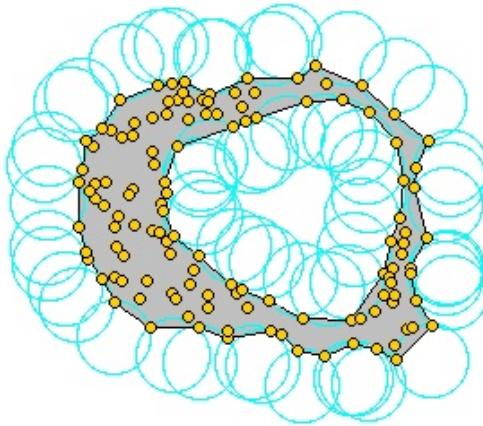


Figure 5.1: Construction of α -shape given a set of points in \mathbb{R}^2 .

with C a constant that can be determined by a simulation and Δ the density of the point set \mathcal{S} defined as:

$$\Delta = \frac{N}{\text{surface area}} , \quad (5.1.2)$$

where N is the number of points in \mathcal{S} and the surface area is the area inside the boundaries of the region formed by the points cloud. Hence Δ is a constant. As mentioned above to find the α -shape of a point cloud we need a triangulation and a possible choise could be the Delaunay triangulation. As explained in [15] we can see a Delaunay triangulation as the dual of a Voronoi diagram. Let us define a Voronoi diagram in a metric space.

Definition 5.1.1. Let X be a space endowed with a distance d and $\mathcal{S} = \{S_1, \dots, S_n\}$ a set formed by subsets of X . The Voronoi cell R_k associated with the set S_k where $k \in \{1, \dots, n\}$ is defined as follows:

$$R_k = \{\mathbf{x} \in X \mid d(\mathbf{x}, S_k) \leq d(\mathbf{x}, S_j) \quad \forall j \neq k\} , \quad (5.1.3)$$

where $d(x, A) = \inf\{d(x, a) \mid a \in A\}$. The Voronoi diagram is defined as the tuple of the cells $(R_k)_{k \in \{1, \dots, n\}}$ that are assumed to be disjoint.

The simplest case that we can have is the two-dimensional case that is the case where $X = \mathbb{R}^2$. The tuple $\mathcal{S} = \{1, \dots, n\} \subset \mathbb{R}^2$ is now a set of points. The Voronoi diagram of \mathcal{S} is a subsection of \mathbb{R}^2 such that every other region around a point $p \in \mathcal{S}$ contains all points that are closer to p than to every point in \mathcal{S} . A triangulation of the point set \mathcal{S} is a set of edges \mathcal{E} whose extremes are points of \mathcal{S} such that the faces of each triangle are bounded by three edges and any edge that is not in \mathcal{E} intersects one of the existing edges. The Delaunay triangulation is the dual graph of the Voronoi diagram: it consists of vertices (the points in \mathcal{S}) and it has an edge between two vertices if the two corresponding faces share an edge.

The Delaunay triangulation triangulates the convex hull of the point set \mathcal{S} . Instead,

the α -shape of a point set is formed only by the triangles (taken from the Delaunay triangulation) that satisfy the " α -test" and therefore is a suitable method to reconstruct the surface formed by a point cloud. Even if α -shapes are a powerful tool to reconstruct surfaces, some simulations show that there exist surfaces that are not described well by α -shapes. Indeed for some particular surface there exist no value of α that includes all desired triangles and deletes all undesired triangles. For instance, since the parameter α depends on the density of the point cloud, is intuitively clear that using α -shapes for a non-uniform points set we won't get a good approximation of the surface. Furthermore, the α -shape method doesn't work well when there is a sharp turn or a joint. In this case α -shapes often give a "webbed-foot" appearance at such joints since they improperly connect the adjacent surfaces. Hence a generalization of "classical" α -shapes is required. In the next section a method to solve the "density problem" for two separated and close objects is described. In [16] Teichmann and Capps present "Density-scaled α -shapes". The first step of this method is to make a triangulation of the point cloud. Then the key idea is to compute somehow the point-density of each point and use this to get an approximation of the point density of a triangle. In this way one can reduce the α -value in areas where the triangle's point density (see equation 5.1.6 for the definition) is higher than average in such a way that is possible to obtain a finer level of detail for areas that have an higher density. More precisely, each point $\mathbf{p} \in \mathcal{S}$ has a local point density defined as

$$\delta(\mathbf{p}) = \sum_{\mathbf{q} \in \mathcal{S}} \left(1 - \frac{d(q, p)}{\lambda}\right) \quad \forall \mathbf{q} \text{ such that } d(\mathbf{p}, \mathbf{q}) < \lambda, \quad (5.1.4)$$

where λ is the constant radius of the local neighborhood and $d(\mathbf{x}, \mathbf{y})$ is the Euclidean distance. When local density is larger than the average, that is when

$$\delta(\mathbf{p}) > \frac{1}{|\mathcal{S}|} \sum_{\mathbf{q} \in \mathcal{S}} \delta(\mathbf{q}) \quad (5.1.5)$$

we know some properties about the region surrounding \mathbf{p} . For instance, if the point set is uniformly distributed then it is possible to find areas with a high-density in the case where there are two closely separated surfaces. In point sets of non-uniform distribution, high densities are found when the surface presents a joint discontinuity. The algorithm developed by Teichmann and Capps is structured as follow. After computing density information for each point they make a triangulation of the point set. Then they calculate the average density $\delta(t)$ for each triangle Δ_{abc} defined as:

$$\delta(t) = \frac{\delta(a) + \delta(b) + \delta(c)}{3\mu}, \quad (5.1.6)$$

where μ is the global average density of the entire point set \mathcal{S} . If $\delta(t)$ is greater than 1 the density of the point cloud is higher. Hence is necessary to define another value of α :

$$\alpha' = \frac{\alpha}{\delta(t)^\sigma} \quad (5.1.7)$$

where σ is a value that is adjusted by the user. If δ is less than 1 the α -value is not modified. In this way it is possible to have a finer precision on the shape formed by the point set where the density is higher than the average density. Hence it is possible to distinguish two separated objects with different density.

- 5.2 The two-faceted cup**
- 5.3 Results for a TIR collimator**
- 5.4 The triangulation refinement approach**
- 5.5 The two-faceted cup**
- 5.6 Results for a TIR collimator**
- 5.7 Results for a Parabolic reflector**
- 5.8 Results for the Compound Parabolic Concentrator (CPC)**

Chapter 6

The inverse ray mapping method: analytic approach

6.1 Explanation of the method

6.2 The two-faceted cup

6.3 The multi faceted cup

For two-dimensional systems every ray in the PS of a line is given by a two-tuple point. Therefore, the PS of every line is a two-dimensional space. The position coordinate in the PS of line i is the x -coordinate of the intersection point between the ray and the line i . The direction coordinate is the sine of the angle that the ray forms with respect to the normal of the line i multiplied by the index of refraction of the medium in which the ray is located. Let's now introduce some notation before explaining the details of the method. We indicate the PS with $S = Q \times P$, where Q is the set of the position coordinates q and P is the set of the direction coordinates $p = n\sin\tau$ with τ the angle between the ray and the normal ν of the line and n is the index of refraction of the medium in which the line is located.

6.4 Results for the two-faceted cup

6.5 Results for the multi-faceted cup

6.6 Discussions

Chapter 7

The extended ray mapping method

- 7.1 Explanation of the method
- 7.2 Bisection procedure
- 7.3 Results for a parabolic reflector
- 7.4 Results for two different kind of TIR-collimators

Chapter 8

Extended ray mapping method to systems with Fresnel reflection

Chapter 9

Discussion and conclusions

Summary

In this thesis we studied the light propagation within optical systems. Optical engineers are interested in design systems in such a way the desired output distribution is obtained. The goal in illumination optics is to obtain the desired output distribution of light. To this purpose the ray tracing procedure is widely used. Ray tracing is a forward method where a set of rays is traced within the system from the source to the target. The propagation of light is determined computing the position and the direction of every ray for all the optical surfaces that it encounters. There are many ways to implement the ray tracing process. Monte Carlo (MC) ray tracing is often used in non-imaging optics. Rays are randomly traced from the source to the target and each time that a ray hits an optical surface the coordinates of the intersection point of the ray with the surface and the new ray direction are calculated. The output variables are computed dividing the target into intervals, the so-called bins, and counting the rays that fall into each bin. To obtain the desired accuracy, millions of rays are required, therefore the method is extremely computationally expensive and it converges as the inverse of the square root of the number of rays traced.

MC ray tracing can be improved using as sample of points a low discrepancy sequence instead of random points. Discrepancy can be interpreted as a measure of how much the sample distribution differs from a uniformly distributed sample. The discrepancy is therefore zero for uniformly distributed points. A low discrepancy sequence gives a sample of points which are regularly distributed but not exactly uniform distributed. Quasi Monte Carlo (QMC) method considers this kind of sequences as sample of points. Therefore, QMC ray tracing is implemented tracing a set of rays whose position and direction are given by the coordinates of a low discrepancy sequence of points. The main advantage of QMC method is its rate of convergence, it is faster than MC for low dimensional problems. Nevertheless, it has some disadvantages. First, it is not easy to give an error estimation for QMC method. Second, for high dimensional spaces the QMC can become very slow. Third, it is still a binning procedure. Hence, the accuracy depends both on the number of rays traced and on the umber of bins.

In order to improve the existing methods the phase space (PS) of the optical system is considered in this thesis. The PS of an optical surface gives information about the position and the direction of every ray on that surface where the direction is expressed with respect to the normal of the surface. In PS, the ray's direction is given by the sine of the angle that the ray forms with respect to the normal of the surface multiplied by the index of refraction of the medium in which the ray is located. In two dimensions, the PS is a two-dimensional space where the coordinates of every

ray are specified by one position coordinate and one angular coordinate. For three dimensional systems the PS is a four dimensional space because every ray is specified by two position and two angular coordinates. Our idea is to use the structure of PS to trace only the rays close to the discontinuities of the luminance at the target PS. Two new approaches based on PS are presented in this work. They are tested for two-dimensional systems.

The first method is called ray tracing on PS and it is based on the source and the target PS representation of the optical system. It takes into account the sequence of optical lines that each ray hits when it propagates inside the system, that is the ray path. We note that the source and target phase spaces are partitioned into different regions each of them is formed by the rays that follow the same path. The idea is to use the edge-ray principle proved by Ries and Rabl (1994) which states that the area of these regions is conserved: all rays that are neighbors at the source phase space remain close to each other at the target. To this purpose, a nonuniform triangulation of the source PS is constructed in such a way that new triangles are added to the triangulation only where boundaries occur. Assuming constant brightness, we only need to compute the boundaries of the regions in target PS to obtain the output photometric variables. The details of PS ray tracing are explained in Chapters 3 and 4 where we test the method for optical systems where both reflection and refraction laws are involved. Numerical results are presented in Chapter 5 where we show that ray tracing on PS is faster and more accurate compared to MC ray tracing.

The second method employs not only the source and the target PS, but also the PS of *all* the other lines that constitute the system. All lines can be modeled as detectors of the incident light and emitters of the reflected light. Moreover, we assume that the source can only emit light and the target can only receive light. Therefore, one PS is taken into account for the source and one for the target. For the other surfaces both the source and target PS are considered. Furthermore, instead of starting from the source, the new method starts tracing back rays from target PS. In order to determine the coordinates of these rays, an inverse map from the target to the source PS is constructed as a concatenation of the maps that relate the PS of two different lines. Employing this map we are able to detect the rays that in target PS are located on the boundaries of the regions with positive luminance. In Chapter 6 we explain how to implement the method for systems formed by straight and reflective lines. In this particular case, the boundaries of the regions that form every PS can be computed analytically. This allows us to obtain an analytic target intensity distribution. The results are shown for a two-faceted cup and a multi-faceted cup. In both cases we note significant advantages both in terms of the accuracy and the computational time. Then, the method is also developed for systems formed by curved lines. In this case the boundaries cannot be determined analytically and therefore a numerical procedure is involved. In particular, we apply a bisection method on target PS. Also in this case we compare our method to MC ray tracing and we observe significant advantages using the PS method. The results for systems formed by curved lines where reflection and refraction law are involved are shown in Chapter 7. Finally, the ray mapping method in PS is applied to systems where also Fresnel reflection is taken into account. The explanation of the procedure and the numerical results are given in the last chapter of this work.

List of publications

Refereed journal papers:

1. C. Filosa, J. ten Thije Boonkkamp, W. IJzerman, *Ray tracing method in phase space for two-dimensional optical systems*, Applied optics, Optical Society of America, Volume 55, pp 3599–3606, 2016.
2. C. Filosa, J. ten Thije Boonkkamp, W. IJzerman, *Phase space ray tracing for a two-dimensional parabolic reflector*, Mathematics and Statistics, Volume 5, pp. 135 - 142, 2017.

Non refereed proceeding papers:

1. C. Filosa, J. ten Thije Boonkkamp, W. IJzerman, *Inverse ray mapping on phase space for two-dimensional optical systems*, Optical Fabrication and Testing, OSA, 2017.

Curriculum Vitae

Carmela Filosa was born on November 28, 1985 in Torre del greco, Italy. She finished the high school in 2003 at Liceo Scientifico Statale "G. Marconi", Colleferro. She obtained her bachelor (2008) and Master (2013) degree in Mathematics at the University of Rome "La Sapienza", Italy. In March 2014 she moved in Eindhoven (the Netherlands) to start a PhD project at the Eindhoven University of Technology in the department of Mathematics and Computer Science. The PhD project was under the supervision of Wilbert IJzerman and Jan ten Thije Boonkkamp. The research conducted in her doctoral studies was funded by STW. The daily work took place at the Centre for Analysis, Scientific computing and Applications (CASA) of TU/e and sometimes at the department of Philips Lighting of the High Tech Campus in Eindhoven. The results of her research are presented in this thesis.

Acknowledgments

Bibliography

- [1] E. F. Zalewski, “Radiometry and photometry,” *Handbook of optics*, vol. 2, pp. 24–1, 1995.
- [2] J. Chaves, *Introduction to nonimaging optics*. CRC press, 2015.
- [3] “Luminous efficacy-wikipedia the free encyclopedia,” https://commons.wikimedia.org/wiki/File:CIE_1931_Luminosity.png media/File:CIE 1931 Luminosity.png.
- [4] A. V. Arecchi, R. J. Koshel, and T. Messadi, “Field guide to illumination,” SPIE, 2007.
- [5] H. Zhu and P. Blackborow, “Etendue and optical throughput calculations,” *En-ergetiq Technology, Inc., Woburn, MA*, 2011.
- [6] R. P. Feynman, “Feynman lectures on physics. volume 2: Mainly electromagnetism and matter,” *Reading, Ma.: Addison-Wesley, 1964, edited by Feynman, Richard P.; Leighton, Robert B.; Sands, Matthew*, 1964.
- [7] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.
- [8] E. Hecht, *Optics*. Parson Addison-Wesley, 2002.
- [9] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics, Vol. I: The new millennium edition: mainly mechanics, radiation, and heat*, vol. 1. Basic books, 2011.
- [10] P. H. Jones, O. M. Maragò, and G. Volpe, *Optical tweezers: Principles and applications*. Cambridge University Press, 2015.
- [11] D. Z. Ting and T. C. McGill Jr, “Monte Carlo simulation of light-emitting diode light-extraction characteristics,” *Optical Engineering*, vol. 34, no. 12, pp. 3545–3553, 1995.
- [12] K. B. Wolf, *Geometric optics on phase space*. Springer Science & Business Media, 2004.
- [13] H. Ries and A. Rabl, “Edge-ray principle of nonimaging optics,” *JOSA A*, vol. 11, no. 10, pp. 2627–2632, 1994.

- [14] H. Edelsbrunner and E. P. Mücke, “Three-dimensional alpha shapes,” *ACM Transactions on Graphics (TOG)*, vol. 13, no. 1, pp. 43–72, 1994.
- [15] J. Portegies and P. Lighting, “Fast ray tracing in phase space for optical design,” 2013.
- [16] M. Teichmann and M. Capps, “Surface reconstruction with anisotropic density-scaled alpha shapes,” in *Visualization’98. Proceedings*, pp. 67–72, IEEE, 1998.