UNIVERSITY OF
KWAZULU-NATAL

INYUVESI
YAKWAZULU-NATALI

Measurement and Modelling of the Directional Scattering of Light

Derek John Griffith

2015

Measurement and Modelling of the Directional Scattering of Light

by

Derek John Griffith (212561361)

BSc (Physics)

Submitted in partial fulfilment of the requirements

for the degree of Master of Science in Physics

School of Chemistry and Physics

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

27 February 2015

Declaration

This thesis describes the work undertaken at the University of KwaZulu-Natal under the supervision of Dr N. Chetty between July 2012 and February 2015.

I declare the work reported herein to be my own research, unless specifically indicated to the contrary in the text.

Signed: .....

D. J. Griffith

On this ..... day of 2015

I hereby certify that this statement is correct.

Signed: .....

Dr N. Chetty (Supervisor)

On this ..... day of 2015

Acknowledgements

My wife, Ingrid Swart, has made this project possible in many more ways than she knows. I dedicate the result to her.

My supervisor, Naven Chetty, has played a crucial role not only in providing guidance on all fronts and keeping this work moving ahead, but also in creating the original opportunity.

I thank also my employer, CSIR, and all of my colleagues there for providing an environment in which further study is encouraged and rewarded.

Abstract

The quantum nature of light suggests that a photon can interact with matter in two primary ways. Firstly and perhaps more simply, the photon could be absorbed or secondly and more complex, it could be scattered into a new direction of propagation. The scattering process can be thought of as probabilistic, with a statistical distribution of possible new directions of travel with respect to the original. In the case of interaction with a small particle of matter, the probability distribution is referred to as the *phase function*. In the case of scattering at a surface interface between two bulk materials, the new direction of travel is distributed according to a function called the *Bidirectional Scattering Distribution Function* (BSDF). The BSDF depends on both the direction of arrival and the direction of scatter (hence *bidirectional*), the type of material and the condition of the surface as well as the wavelength of light.

This work explores a number of areas related to the BSDF, with special attention to the effects of random light scatter in high performance optical imaging systems such as space telescopes. These demanding imaging applications require optical components manufactured to very high standards with respect to shape, smoothness and cleanliness. This means that random scatter from the surfaces of these optical components must be controlled to very low levels. The measurement of very weak optical surface scatter is therefore a problem of particular interest. An interferometric technique has been proposed here for improving the quality of such measurements. The interference effects produced in the image by this technique were analysed using Nijboer-Zernike diffraction theory, leading to a journal publication in Current Applied Physics.

Contents

Table of Abbreviations

Abbreviation	Definition
ACV	Auto-Covariance function
AFM	Atomic Force Microscope (or Microscopy)
ARS	Angular Resolved Scattering function
BK	Beckmann-Kirchhoff (Scattering theory)
BRDF	Bidirectional Reflectance Distribution Function
BSDF	Bidirectional Scattering Distribution Function
BTDF	Bidirectional Transmission Distribution Function
CNZ	Classical Nijboer-Zernike (Theory)
DUV	Deep Ultraviolet
EO	Earth Observation
ENZ	Extended Nijboer-Zernike (Approach or Analysis)
EUV	Extreme Ultraviolet
FT	Fourier Transform
FCT	Fourier Cosine Transform
GHS	Generalised Harvey-Shack (Scattering theory)
HS	Harvey-Shack (Scattering theory)
IC	Integrated Circuit

MTF	Modulation Transfer Function
NA	Numerical Aperture
NSC	Non-Sequential Component (Zemax®)
OPD	Optical Path Difference
OPL	Optical Path Length
OTF	Optical Transfer Function
PSD	Power Spectral Density
PSF	Point Spread Function
PTF	Phase Transfer Function
RR	Rayleigh-Rice (Scattering theory)
SRF	Spectral Response Function
TIS	Total Integrated Scatter
UV	Ultraviolet

Chapter 1

Background and Motivation

1.1 Introduction

Scattering of light from material surfaces and particles is an important field of study for numerous scientific and technological disciplines including computer vision/graphics [1–3], radiative transfer [4, 5], optical systems [6, 7] and Earth observation by remote sensing [8–10]. Major determinants of the scattered spatial light field are the type of material and the surface topography (profile, shape, texture, roughness, structure) [11]. The intensity of light scattered in different outgoing directions varies with both the light wavelength and the directions from which the light arrives at the surface or particle [11, 12].

When considering material surfaces, the concept of the Bidirectional Scattering Distribution Function (BSDF [13]) is a useful analytical tool for describing the directional manner in which a material surface scatters electromagnetic (EM) radiation [11]. In this work, the theory, measurement and modelling of the BSDF, particularly as it relates to material surface topography has been investigated. This is of special interest since it describes the effect of undesirable light scattering in optical imaging systems [7]. In

critical applications such as space telescopes, unwanted *stray light* reaching the image sensor reduces the quality, contrast and viability of the image [14]. This applies particularly to imaging systems operating at short wavelengths such as lens systems for ultraviolet lithography, a key enabling technology in the manufacture of high density integrated circuits [15]. Shorter wavelengths are more susceptible to scattering from small topographic surface features as well as small particles.

When considering high performance optical imaging applications that are sensitive to stray light, a few questions need to be answered:

1. Is there a practical and reliable way of computing the BSDF of an optical surface from measurements of the surface topography? It is significantly important to answer this question appropriately to successfully implement a precision imaging system for the following reasons:

   (a) Surface topography is comparatively easy to measure since commercial instruments for surface profilometry (down to atomic scale) are readily available.

   (b) BSDF is comparatively difficult and slow to measure since it must be measured over a large range of incidence and scatter angles as well as over different wavelengths.

   (c) High precision instruments for direct measurement of the BSDF (*scatterometers* [16]) are typically research instruments designed, implemented and/or operated by metrological or academic institutes [17].

   (d) It is much more practical and cost-effective to specify surface topography statistics rather than BSDF when optical surfaces are manufactured,

   (e) When undertaking computational analysis of the performance of a complete optical imaging system, it is the surface BSDFs that are required.

2. Are there any ways in which scatterometers can be improved such that direct BSDF measurements of optical surfaces or components could become more routine and accessible? Resolving this question is important because:

   (a) Surface topography is not the only factor contributing to unwanted scatter from optical components. Other contributors include sub-surface damage, optical coating structure and internal (bulk material) scatter such as from bubbles and artefacts within glass components, which implies that scatter produced through these other mechanisms is visible to scatterometers but not via surface topography measurements.

   (b) Direct BSDF measurements are required for validation of the models and methods used for computing BSDF from surface topography (see preceding question).

3. What are factors that can degrade the performance of an optical system that is sensitive to stray light? It is important to know and understand all of the possible mechanisms by which scattered stray light can originate in precision optical systems as well as other factors that can affect performance.

4. What are the most advanced current methods for predicting and measuring the performance of critical, complex imaging systems which are very sensitive to both optical aberrations (essentially geometric errors) and stray light? This question is highly relevant to state-of-the-art, short wavelength imaging systems such as those used in deep and extreme ultraviolet (DUV and EUV) lithography.

The main objective of this work is therefore to explore possible answers to the above questions.

1.2 Precision Imaging Optical Systems

Of particular interest in precision optical imaging systems is the BSDF of the lens and mirror surfaces. Precision imaging optics have a broad range of applications such as microscopy, imaging astronomy, Earth Observation (EO) from space and lithography [15, 18–20]. These systems normally comprise many optical surfaces and components and thus the effect of unwanted scatter from these surfaces is cumulative (see Figure 4.1 for an example of a lithography lens system).

Lithography is a particularly demanding imaging technique applied in the manufacture of state-of-the-art electronic Integrated Circuits (IC) and thus the effect of unwanted scatter must be well-known and accounted for. In fact, in any of the above applications, undesirable scattered light from the optical surfaces, within the bulk of the optical components or from mechanical components, could reach the image plane. Should such scattered or *stray* light reach the image plane, the quality (contrast, dynamic range) and utility of the images produced by these optical systems is significantly reduced [7]. With ever increasing demand for better image quality at short wavelengths, such as in extreme ultraviolet (EUV) lithography [20], it becomes increasingly necessary to have powerful techniques for measuring weak scatter from single optical surfaces as well as the cumulative effect of weak scatter in imaging systems.

The instruments used to measure scatter from individual optical surfaces or within complete optical systems are referred to as *scatterometers* [11, 16]. When considering precision optical surfaces, there is an implicit assumption that upon interaction with the surface, it is possible to differentiate the departing radiation into scattered and unscattered components. Scattered radiation must be very weak to maintain high image contrast. The unscattered component, also referred to as the *specular* component is much more intense in this situation and thus results in a clearer image. A scatterometer is therefore extremely useful in distinguishing between these two components [11] and

thus must have very high measurement dynamic ranges, in order to discern weak scatter departing from the surface in directions very close to the direction of the specular beam [21].

The work presented herein began as an effort to refine the techniques for addressing the problems in measuring weak scatter near the direction of the specular beam. However, it soon became clear that the classification of light into scattered and unscattered components is essentially arbitrary. In reality, there is a continuum of scattering into wide angles by small surface topographic features or particles through to narrow angle scatter from the limiting aperture of the optical beam (diffraction). Nevertheless, there is a need to assume that there is a measurable distinction between *random* or *incoherent* scattered light and the *specular* or *coherent* component [21]. The random/incoherent scattered component may also be referred to as *diffusely* scattered light [11].

Due to diffraction [22], the limiting aperture of an imaging system, together with the wavelengths of light utilised by the system determine the limit of fine detail that can be discerned in the image.

Imaging optical systems can be idealised as linear and shift-invariant systems [23] for many purposes and can thus be attributed with a transfer function [24, 25], called the Optical Transfer Function (OTF). The Fourier Transform (FT) of the OTF is the system Point Spread Function (PSF). The PSF expresses the spatial intensity of the image of a perfect point of light taken as the object. The PSF and the OTF are an FT pair that both describe the quality of the image.

The above problem can thus be reduced to a requirement for performing very high dynamic range measurements of the PSF/OTF of an imaging system.

The following chapters consider the measurement and applications of the BSDF concept. Thereafter, the measurement and specification of scattering from optical surfaces is discussed in the context of performing high dynamic range measurements of the

PSF. This finally leads to the problem of using such PSF measurements to extract lens error (aberration) diagnostic information from the precision PSF measurements. The latter task is referred to as *aberration retrieval*.

1.2.1 Aberration Retrieval

Traditionally, laser interferometry has been a powerful technique for performing aberration retrieval [26,27]. Interferometry can be used to measure the shape of the wavefront emerging from the exit pupil of the imaging system under test. The emergent wavefront shape is quantified as the *pupil function* [28]. The pupil function is also related to the PSF in that the PSF is the Fourier transform of the autocorrelation of the pupil function (see Figure 2.2).

However, conventional interferometry for evaluation of optical system performance suffers from a number of drawbacks:

1. Usage of a laser as a diffraction-limited point source restricts measurement to a single wavelength. Many optical systems of interest are *polychromatic* in that they respond to a range of wavelengths and the PSF/OTF always varies with wavelength [24].

2. Long coherence length lasers are generally required to obtain decent interferogram contrast with large path differences. Such lasers inherently produce noisy images due to laser speckle phenomenon [27].

3. Interferometers tend to produce spurious fringe patterns from stray (ghost) reflections in the system and from dust particles or defects on the lens/mirror surfaces.

4. The spatial resolution of the camera used to record the interferogram limits the range of spatial frequencies represented in the measurement [27, 29].

5. The PSF is only indirectly measured by performing an autocorrelation of the pupil function computationally reconstructed from the interferogram [27].

Spurious fringes and laser speckle increase the noise in interferometric measurements, reducing the overall signal-to-noise ratio of the measurement result [27] and hence increasing the uncertainty.

The capability to measure the PSF directly and also perform diagnostic aberration retrieval is an attractive alternative proposition. Unfortunately, this is inherently an ill-posed problem because autocorrelation is generally an irreversible operation. The pupil function carries aberration information and the PSF is the Fourier transform of the autocorrelation of the pupil function. However, there is a growing body of knowledge which makes use of intensity measurements in the vicinity of the focal plane to overcome this barrier. Of special interest in this regard is the Extended Nijboer-Zernike (ENZ) approach which allows for accurate computation of the EM field in the focal plane region. Aberration retrieval is therefore facilitated by leveraging the measurement diversity made accessible by the ENZ approach.

The journal article, submitted to Current Applied Physics and which is reproduced in Chapter 6, is the centrepiece of this MSc. It arises *inter alia* from consideration of an additional or alternative method for achieving measurement diversity in order to make use of the ENZ approach for aberration retrieval.

Chapter 2

Theoretical Background

2.1 Introduction

In this chapter, relevant definitions, theory and models for the BSDF are presented. The first question posed in the introduction to Chapter 1 (Is there a practical and reliable way of computing the BSDF of an optical surface from measurements of the surface topography?) is explored and further discussed.

In addition to seeking an answer to this question, a generic method is presented for using such predicted BSDFs in stray light analysis of precision optical imaging systems.

2.2 Definitions and Terminology

Much of the formal terminology for scattering was established by Nicodemus *et. al* [13] at the National Institute of Standards and Technology (NIST) in 1977. Details of the definitions and notation may vary amongst the various disciplines making use of these or similar concepts. In part, this is due to large differences in how these concepts are applied amongst these disciplines and because the literature has developed independently, to some extent, in each application area.

2.2.1 Bidirectional Scattering Distribution Function

The Bidirectional Scattering Distribution Function (BSDF) is defined [13] as the differential spectral radiance, dL_λ departing (scattered) from the surface in a particular direction (θ_s, ϕ_s) divided by the differential spectral irradiance, dE_λ arriving at the surface in a particular direction (θ_i, ϕ_i). The directions of arrival and departure are defined using the polar angle relative to the surface normal and the azimuthal angle relative to a specified azimuthal reference,

$$BSDF(\theta_i, \phi_i, \theta_s, \phi_s; \lambda) = \frac{dL_\lambda(\theta_s, \phi_s)}{dE_\lambda(\theta_i, \phi_i)}. \tag{2.1}$$

In general, the BSDF is a function of the directions of both the incident and scattered light directions (hence "bidirectional") as well as the electromagnetic wavelength (λ) dependence.

The BSDF of a material is not directly measurable, since light sources and detectors in reality have finite angular apertures or linear extent. Therefore the measured quantity is better described as the Biconical Scattering Distribution Function (BCSF) [13, 30]. This is a common feature of many kinds of measurements in which a physical quantity is not measured instantaneously but sampled over a finite spatial or temporal interval, often in a non-uniform manner [29, 31]. This weighted sampling, performed by the measuring instrument is quantified by the *apodization* or *instrument* function [32]. The *instrument* function can be defined as the Fourier Cosine Transform (FCT) of the *apodization* function.

Apodization functions are referred to by other names in different contexts or disciplines. For example, when describing the relative spectral response of an optical detector, the spectral apodization function may be called a Spectral Response Function (SRF) or Relative Spectral Response (RSR). In signal processing, the apodization function may

be called a *window function* or *tapering function* [33].

In this respect, BSDF is not fundamentally different to other physical measurands such as optical power spectra which are thought to have a "true" underlying value, which is imperfectly measured [34]. Accepting that the BSDF is not directly measurable, when referring to "BSDF measurements", this should be understood to include any real measurements that are weighted spatial or spectral averages of the "true" BSDF.

2.2.2 BRDF and BTDF

The BSDF is the more general form of the bidirectional scattering function. It includes the possibility of (transmitted) forward scatter into the hemisphere beyond the hemisphere of arrival at the surface, as well as (reflected) backscatter into the hemisphere of arrival. The BSDF can therefore be split into two functions, the Bidirectional Reflectance Distribution Function (BRDF) and the Bidirectional Transmittance Distribution Function (BTDF) [13].

In the literature the formal differences between BSDF, BTDF and BRDF are not always apparent and the terms can be used informally or interchangeably. In the field of Earth observation from space or airborne platforms [30], the directional reflectance of the Earth surface is commonly and correctly referred to as the BRDF. In computer graphics [2], BRDF is the most commonly used term, although transparent or translucent objects may occur.

In relation to optical systems, the BTDF is usually the more relevant for lens surfaces, while the BRDF is more relevant for mirror surfaces for obvious reasons. When it comes to the opaque mechanical parts of a lens system, the BRDF (of black paints for example [35]) is formally correct and almost exclusively the term that is used in practice.

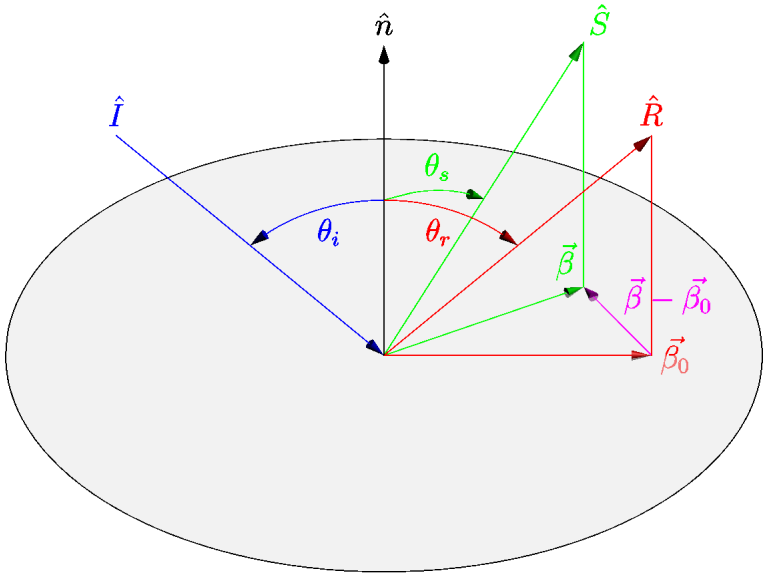

Figure 2.1: Scattering Geometry

2.3 Definitions Specific to Optical Surfaces

When dealing with the optical surfaces of lenses and mirrors, there are some specific geometrical definitions relating to directional scattering that become relevant and that are commonly used in related literature [11].

Figure 2.1 shows a flat, reflective optical surface with unit normal vector $\hat{n}$ pointing away from the surface. The following definitions largely follow Dittman [36] and the Zemax® manual [37]. A plane electromagnetic wave travelling in the direction designated by the unit vector $\hat{I}$ intercepts the surface at an *angle of incidence* θ_i measured relative to the surface normal. The coherent, specular reflection leaves the surface travelling in the direction given by the unit vector $\hat{R}$ at an angle of θ_r relative to the surface normal vector. The unit vectors $\hat{n}$, $\hat{I}$ and $\hat{R}$ all lie in a plane called the *plane of incidence* [22]. The unit vector $\hat{S}$ indicates a possible direction of departure for incoherent, random optical scatter occurring at the surface. Random scatter may occur due to residual surface fabrication roughness, a surface defect such as a scratch, or surface contamination such as a dust particle. An important quantity that will arise often in

the discussions to follow is the sine magnitude of the scatter angle. This is a scalar quantity, designated β, defined as the magnitude of the difference vector between the projections of $\hat{R}$ and $\hat{S}$ onto the plane of the surface. In Figure 2.1, the projection vector of $\hat{R}$ onto the surface plane is labelled $\vec{\beta_0}$ and the projection of $\hat{S}$ is labelled $\vec{\beta}$. The sine magnitude of the scatter angle is defined as:

$$\beta = |\vec{\beta} - \vec{\beta_0}|. \tag{2.2}$$

The Zemax® manual [37] defines $\vec{x} = \vec{\beta} - \vec{\beta_0}$ and $x = |\vec{x}|$, but this notation will be avoided and β will be used here and again in §2.5.1. Note that $\beta = 0$ for scattering in the specular direction and that β can approach a maximum value of 2 for the extreme case of backscattering ($\hat{S} = \hat{I}$) of light at grazing incidence ($\theta_i \to 90°$).

The significance of β is best understood in terms of first order diffraction due to a grating having a frequency of f repetitions (cycles) per unit distance in the plane of the surface. For normal incidence ($\theta_i = 0$), the simple grating equation [22] gives the relationship between f and the first order diffraction scattering angle θ_s as follows:

$$f\lambda = \sin\theta_s, \tag{2.3}$$

where λ is the wavelength of light and β is the generalisation of $\sin\theta_s$ to arbitrary scattering directions and angles of incidence as defined in Equation 2.2. For a particular wavelength, every direction of scatter away from the specular direction can be associated with first order diffraction from a grating on the surface having a particular spatial frequency and azimuthal orientation. β is the link between the angle of scattering, the spatial frequency and the wavelength and it retains the simple property that for all angles of incidence and scattering:

$$\beta = f\lambda. \tag{2.4}$$

The above definition also applies to refractive transmission through an optical surface such as the surface of a lens. In general, the surfaces of lenses and mirrors are curved, so there is an implicit assumption that the BSDF can be specified and measured over an area that is sufficiently small to be treated as locally flat.

In many cases of interest to stray light analysis, the surface BSDF can be expressed as a function of β.

2.3.1 Total Integrated Scatter

The Total Integrated Scatter (TIS) is defined as that proportion of the total radiant power remaining after light interaction with the surface that is scattered out of the specular beam [38]. That is, a certain quantity of radiant power (flux) is incident on the surface, of which some is absorbed and the remainder is reflected or transmitted. The TIS is the ratio of the surviving flux scattered out of the specular beam to the total surviving (non-absorbed) flux. The most elementary situation is the case of an optical mirror, where only the reflected flux is of interest. Reflectance in general is defined as a ratio of reflected to incident flux [13]. If the total reflectance for a particular angle of incidence θ_i is ρ_t, the specular reflectance is ρ_r and the diffusely scattered reflectance is ρ_s then the TIS can be expressed as [38]:

$$TIS = \frac{\rho_s}{\rho_t} = \frac{\rho_s}{\rho_s + \rho_r} = \frac{\rho_t - \rho_r}{\rho_t} = 1 - \frac{\rho_r}{\rho_t}. \tag{2.5}$$

In the context of isotropic mirror surfaces (or if there is interest in reflectance only), the TIS for a particular angle of incidence is also an integral of the incoherent BRDF as:

$$TIS = \iint BRDF(\theta_i, \phi_i, \theta_r, \phi_r), \tag{2.6}$$

or as a function of β as:

$$TIS = \frac{2\pi}{\rho_t} \int BRDF(\beta)\beta d\beta. \tag{2.7}$$

It is very important to note that the BRDF used here does *not* include the specular ("unscattered") beam. In this sense, it does not conform to the definition of BSDF/BRDF given above in Equation 2.1. An obvious problem with the definition of TIS here is that, as pointed out above, the distinction between "scattered" and "unscattered" is rather arbitrary and requires, in addition, some statement about the angle of scatter that is taken as the boundary between scattered and specular. In the case of the simple Harvey scatter model discussed in §2.5.1, the boundary angle is taken as $\beta = 0.01$ (see Equation 2.2 for a definition of β), which for modest incident angles is close to an angle of 0.01 radians from the specular beam. Available literature does not justify defining this particular boundary and it thus has the appearance of being an arbitrary choice.

Nevertheless, TIS is still commonly used and measured [38, 39], and this particular problem is sometimes referred to as "bandwidth-limiting" of the roughness spatial frequencies [40] or the "relevant" spatial frequencies. Upper and lower spatial frequency limits are both relevant. In the case of scatterometer type measuring instruments [16], the boundary angle is usually given as a specification. TIS can be very sensitive to the boundary angle and in certain cases this may bring the usefulness of TIS measurements into question unless the boundary angles are known and reported [38].

2.3.2 Surface Topography and Power Spectral Density

The strong relationship between TIS and surface roughness has been observed and investigated in great detail [11]. Optical scatter in general and TIS in particular has served as a useful means of measuring or characterising surface roughness and topography [38,39,41–43]. The statistical nature of the surface topography is captured using the autocorrelation of the topographic function [11]. The autocorrelation function is the surface topography Power Spectral Density (PSD), which is a function of spatial frequency, f , along the surface [11]. The definition of the 2-dimensional PSD is [44] the square modulus of the Fourier transform of the surface topography function $z(x, y)$,

$$PSD_{2D}(f_x, f_y) = \lim_{L\to\infty} \frac{1}{L^2} \left| \int_0^L \int_0^L z(x, y) \exp[-2\pi(f_x x + f_y y)] \right|^2 , \tag{2.8}$$

where f_x and f_y are spatial frequencies in the x and y directions along the surface. The dimension L is introduced as the length scale over which the topography has been measured. The one-dimensional PSD applies to surfaces of isotropic characteristics calculated by averaging the 2-dimensional PSD over all azimuthal directions along the surface as:

$$PSD(f) = \frac{1}{2\pi} \int_0^{2\pi} PSD(f, \Psi) d\Psi, \tag{2.9}$$

where $f = \sqrt{f_x^2 + f_y^2}$ and Ψ is the azimuthal angle along the surface i.e. $\Psi = \arctan(f_y/f_x)$.

The effective surface roughness for normal incidence and isotropic roughness [38] is:

$$\sigma_{eff} = \left[2\pi \int_{f_{min}}^{f_{max}} PSD(f) f df \right]^{\frac{1}{2}} , \tag{2.10}$$

where the integration limits represent the "relevant" spatial frequencies in any particular situation. In the limit as $f_{min} \to 0$ and $f_{max} \to \infty$, the effective roughness converges

to the total roughness σ. The TIS can be calculated from the effective (or relevant) roughness using the classic relationship [38],

$$TIS = 1 - \exp\left[-\left(\frac{4\pi \cos\theta_i \sigma_{eff}}{\lambda}\right)^2\right], \tag{2.11}$$

which for smooth surfaces can be approximated as:

$$TIS \approx \left(\frac{4\pi \cos\theta_i \sigma_{eff}}{\lambda}\right)^2. \tag{2.12}$$

2.3.3 Autocovariance Function

An alternative but equivalent means of describing surface topography besides the PSD is the surface topography autocovariance function (ACV). For surfaces of zero mean height this is equivalent to the surface topography autocorrelation function. The surface ACV and PSD are a Fourier transform pair whereby ACV is used for the purpose of quantifying the surface characteristic correlation length τ_c. This is the lateral spacing at which the ACV falls to $1/e$ of the maximum value [44]. It is also used as a computational stepping stone to the surface PSD as illustrated in Figure 2.3.

2.3.4 Angular Resolved Scattering

A variant of the BSDF for describing the angular scattering of a surface is the Angular Resolved Scattering (ARS, [38, 44]), which arises in the literature. The ARS is defined as the optical flux component $\Delta\Phi_s$ scattered into a solid angle $\Delta\Omega_s$ in the scattering direction θ_s normalised to the solid angle and incident flux Φ_i i.e.

$$ARS(\theta_s) = \frac{\Delta\Phi_s(\theta_s)}{\Delta\Omega_s \Phi_i} = BSDF(\theta_s)\cos\theta_s = BSDF(\theta_s)\gamma_s. \tag{2.13}$$

2.3.5 Surface PSD Relationship to Image Quality

The relationship between surface topography and image quality is one of special significance to the performance of precision imaging systems.

Precision imaging systems can be conceptualised as a sequence of surfaces (interfaces) between bulk materials which manipulate or modify the electromagnetic wavefront as it traverses the surfaces in turn. For example, the topographical features of each surface are cumulatively imprinted on the wavefront by the process of optical phase delay.

Topographical features on the optical surfaces which are much smaller than the wavelength of light down to atomic scale have little effect, causing low-level background Rayleigh scatter which is fairly isotropic [45]. For larger, more relevant spatial frequencies of such topographic features, the PSD of the surface topography is a useful characterisation as illustrated above. The shape of the wavefront (defined by the complex exit pupil function, P) emerging from the imaging optical system, which subsequently converges to the focal plane, carries the signature of the PSDs of all the surfaces through which it has passed. It is instructive to note that the system Optical Transfer Function (OTF) is the autocorrelation of the pupil function [24].

The optical system PSF is the Fourier transform of the OTF and it is the PSF/OTF Fourier transform pair which encodes the quality of the image. The PSD of the optical surfaces at all relevant spatial frequencies is hence directly linked to the image quality. Figure 2.2 shows the relationships between the complex exit pupil function P, the Amplitude Point Spread Function (APSF) which is the complex amplitude of the PSF, the Modulation Transfer Function (MTF), OTF and Phase Transfer Function (PTF). The symbol $\mathcal{F}$ denotes a Fourier transform relationship. The symbol $\star$ denotes correlation.

In optical systems such as those for deep or extreme UV lithography [15, 19, 20], the wavelength of light is very small and the relevant bandwidth of the surface PSD is relat-

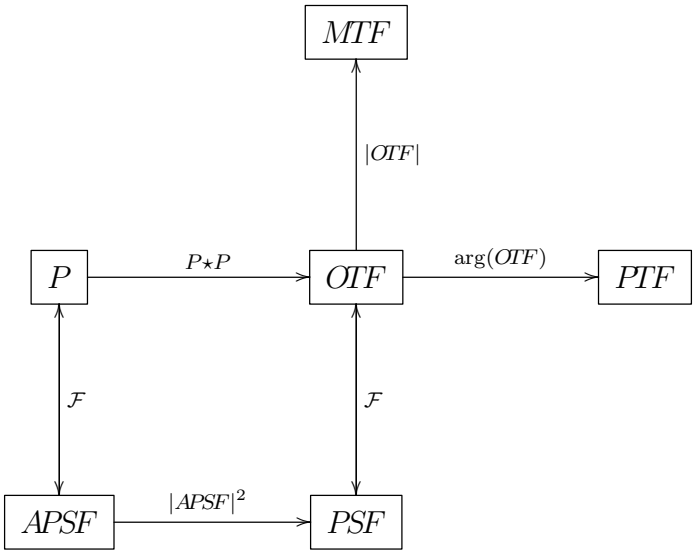

Figure 2.2: Optical Transfer Function Relationships

ively wide, spanning many orders of magnitude [46]. A number of different measurement techniques are required to cover such a broad spatial spectrum, usually incorporating conventional interferometry [26], surface-profiling white-light interferometry [47] and atomic force microscopy (AFM [48]).

Within the realm of precision optical systems, the relationship between surface PSD and BSDF is of particular importance. This is simply because there are many commercial profiling instruments for measuring surface topography down to atomic scale. On the other hand, instruments for precision high dynamic range measurements of surface BSDF are rare and are usually special systems designed and operated by metrological or academic institutes [17, 49–52].

2.4 Optical Scatter

Several theoretical approaches to scatter from moderately rough to nearly smooth surfaces have been published [44, 53]. These approaches are typically developed from Maxwell's equations using various simplifications and approximations [11]. While not

completely rigorous, these theories are very useful both for gaining insight into the scattering phenomenon and for practical purposes. In some cases they also provide a means for solving the inverse problem, namely that of inferring the statistical properties of the surface profile through goniometric or spectro-goniometric scatterometry [39]. Of note are theories known as the classical Beckman-Kirchhoff (BK) theory [54], the Rayleigh-Rice (RR) theory [55], the Harvey-Shack (HS) theory [56] and the most recent Generalised Harvey-Shack (GHS) theory [57, 58]. The RR and GHS theories will be presented briefly to illustrate the relationship between the Power Spectral Density (PSD) of the surface topography and the BSDF.

2.4.1 Rayleigh-Rice Scattering

Also known as the vector perturbation approach, the Rayleigh-Rice theory [11, 55] approaches rigour in the smooth surface limit. It provides a very direct relationship between surface profile and angular scattering and is the cornerstone of statistical smooth surface profiling using goniometric scatterometry. The relationship is as follows:

$$BSDF(\beta) = \frac{4\pi^2 \Delta n^2}{\lambda^4} Q \gamma_i \gamma_s PSD(f). \tag{2.14}$$

The BSDF in this case is a function of the (first order diffraction grating equivalent) amount, β, by which the ray is scattered from the specular direction by a surface spatial frequency component of f cycles per unit distance. β and f are related through the wavelength simply as $\beta = f\lambda$ as explained in §2.3. The change of refractive index between ray arrival and ray departure from the surface is Δn. For reflecting surfaces, $\Delta n = 2$.

Q is a reflectance factor that generally depends on the optical constants of the scat-

tering surface substrate, the scattering geometry and the polarisation state. For small incident and scattering angles, this factor reduces to just the total surface reflectance ρ_t.

The PSD is expressed here as a function of spatial frequency in any direction along the surface which is only possible for isotropic surfaces having no azimuthal dependence of scattering. This is the radial profile of the 2-dimensional PSD. Notice that the BSDF drops off with the cosines of both the polar angles of incidence ($\gamma_i = \cos\theta_i$) and scattering ($\gamma_s = \cos\theta_s$). Note also that since the PSD of a zero mean height distribution typically levels off at a spatial frequency of $f = 0$, the BSDF also levels off at the specular direction where $\beta = 0$. Clearly, this BSDF model *excludes* the specular component.

2.4.2 Generalised Harvey-Shack Scattering

Recently introduced by Krywonos, Harvey and Choi [57, 59], the Generalised Harvey-Shack (GHS) theory is a scalar treatment using the Helmholtz equation [28] and linear systems theory to describe scattering from smooth through to moderately rough surfaces. The theory is valid for arbitrary forms of the PSD (but with Gaussian topographic height statistical distribution) and large angles of incidence and scattering. They introduced a surface scattering transfer function [57, 59]:

$$H_s(\hat{x}, \hat{y}, \gamma_i, \gamma_s) = \exp\left\{ -\left[2\pi\hat{\sigma}_{eff}(\gamma_i + \gamma_s)\right]^2 \left[1 - \frac{ACV(\hat{x}, \hat{y})}{\sigma^2}\right]\right\}, \tag{2.15}$$

where $\hat{x}$, $\hat{y}$ and $\hat{\sigma}_{eff}$ are wavelength normalised coordinates and effective surface roughness. The ARS then takes the form of a correctly scaled Fourier transform of the surface

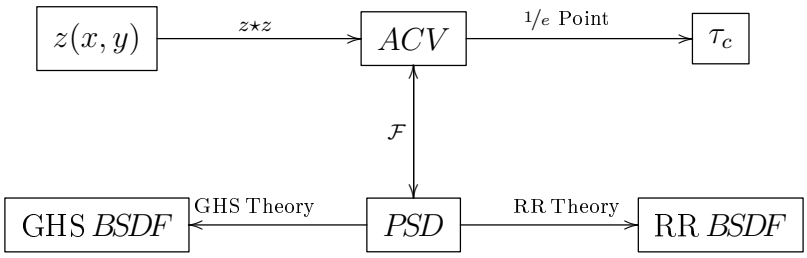

Figure 2.3: Surface Function Relationships

scattering transfer function as:

$$ARS(\theta_s) = Q\gamma_s \mathcal{F}\left\{H_s(\hat{x}, \hat{y}, \gamma_i, \gamma_s)\right\}. \tag{2.16}$$

The reflectance factor Q is defined in the same manner as that used in the RR relationship in Equation 2.14. $\mathcal{F}$ is again the Fourier transform operator.

The GHS approach combines the advantages of the RR and BK approaches while avoiding the limitations of either of these theories. In addition, the theory provides a practical, if computationally expensive means of translating surface topographic profile measurements into an ARS/BSDF model for analysis of stray light effects in imaging systems.

The general relationship between the surface topography function $z(x, y)$, the surface *ACV*, the surface *PSD* and the surface *BSDF* is illustrated in Figure 2.3. The symbol $\star$ denotes the correlation operation.

2.5 Scatter Models for Stray Light Analysis

There are a number of direct models available for scatter resulting from residual fabrication roughness on otherwise smooth optical surfaces. The models that will be discussed below are the 2-parameter and 3-parameter Harvey models [60] and the 3-parameter K-correlation model [11, 36, 61].

2.5.1 The Harvey Models

The 2-parameter Harvey model [60] is a simple inverse power law relating the scatter angle sine magnitude β to the BSDF as follows:

$$BSDF(\beta; b, s) = b(100\beta)^{-s}. \tag{2.17}$$

The first parameter, b is the value of the BSDF at a scatter sine angle magnitude of $\beta = 0.01$, which is near enough to a scatter angle of 0.01 rad measured from the specular direction. For optical surfaces, typical values of b are $0.01 < b < 1$. The second parameter, s is the log-slope of the BSDF decline with respect to β which for optical surfaces lies typically in the range of $1 < s < 3$. This simple model appears to be fractal in the sense that the RR scatter theory predicts a surface radial PSD (see Equation 2.14) that is also a simple inverse power law with the same log-slope across all spatial frequencies. This model is problematic near the specular direction, as it increases without limit as $\beta \to 0$. This behaviour is open to interpretation. It could mean that the RR PSD also increases without limit as $f \to 0$, or that the geometrical specular part of the BSDF is included in the model. The first interpretation is not consistent with smooth optical surfaces and the second interpretation is not useful. For this and other reasons, the 2-parameter Harvey model presents a consistency problem for the definition of the TIS. Although it has proven very useful in stray light analysis, is should be avoided in favour of the 3-parameter Harvey model or the K-correlation model.

For real optical surfaces, the incoherent BSDF levels off at the specular direction and the 3-parameter Harvey model provides this behaviour,

$$BSDF(\beta; b_0, l, s) = b_0 \left[1 + \left(\frac{\beta}{l}\right)^2\right]^{-s/2}. \tag{2.18}$$

This 3-parameter model levels off to a value of b_0 in the specular direction as $\beta \to 0$. At the "shoulder" value of $\beta = l$, the BSDF starts tending towards the simple inverse power law behaviour of the 2-parameter model, also exhibiting the log-log slope of s.

2.5.2 The K-Correlation Model

As one of the available optical surface scatter models in the Zemax® optical design and analysis code [37,62], the K-correlation model warrants special attention. Used by Church [61, 63] and Stover [11] to describe scatter from optical surfaces, the practical use of the model (also known as the ABC model) has been facilitated by Dittman [36] and Gangadhara [64] for the purposes of stray light modelling in imaging systems. The K-correlation 1-dimensional PSD is parametrised [36] as,

$$PSD_{1D}(f) = \frac{A}{[1 + B^2 f^2]^{\frac{s-1}{2}}}, \tag{2.19}$$

and the profile of the 2-dimensional PSD as,

$$PSD_{2D}(f) = \frac{ABg}{[1 + B^2 f]^{\frac{s}{2}}}, \tag{2.20}$$

where,

$$g = \frac{\Gamma(s/2)}{2\sqrt{\pi}\Gamma(\frac{s-1}{2})}. \tag{2.21}$$

$\Gamma(x)$ is the Gamma function [65] and s is the mid-frequency log-slope of the PSD as for the Harvey models. The parameter B is related to the characteristic surface wavelength (correlation distance τ_c) of the surface irregularities [38] and A is a normalisation factor. The K-correlation model provides that the PSD and hence also the BSDF follow an inverse power law over the mid-section of the spatial spectrum, described as "fractal" and commonly the result of subjecting glassy materials to conventional optical (grind

and polish) manufacturing methods [61,66].

Dittman [36] provides explicit expressions for A by normalising the PSD to the total or effective surface roughness, σ using the RR relationship to BSDF and from there proceeds to analytical expressions for the BSDF (not reproduced here).

A convenient feature of the K-correlation model is that the ACV (the Fourier transform of the PSD) which appears in the GHS expression (Equation 2.15) can be calculated analytically from the K-correlation fit parameters (A, B and s) using [67]:

$$ACV(r) = \sqrt{2\pi}\frac{A}{B}\frac{2^{(1-s)/2}}{\Gamma\left((s-1)/2\right)}\left(\frac{2\pi r}{B}\right)^{(s-2)/2}\mathcal{K}_{(s-2)/2}\left(\frac{2\pi r}{B}\right), \tag{2.22}$$

where $\mathcal{K}_n(x)$ is the modified Bessel function of the second kind [65] and order n.

2.5.3 Scatter Model Selection

The choice of scatter model depends on the accuracy with which the scatter must be accounted for in the application. For critical analysis of stray light effects in imaging systems, the K-correlation model is preferred. However, model selection also depends on the data type that is available for compilation/analysis of the model. The 2-parameter Harvey model provides unrealistic behaviour near the specular direction and should be avoided in favour of the 3-parameter Harvey model if the small-angle roll-off of the BSDF is known or can be determined. Likewise, if information about the large angle roll-off is also available, then the K-correlation model should be used in favour of the 3-parameter Harvey model. Selection of model can be assisted by checking the sensitivity of analysis results to the input K-correlation parameters.

2.6 Surface Profilometry and BSDF

There is a strong need in the optical modelling community, especially in precision imaging systems, to have the capability to synthesise a reliable surface BSDF model from surface profile (topography) measurements [61,68]. Such a BSDF model allows for reliable stray light analysis of precision imaging optical systems without having to perform direct measurements of the surface BSDF [69]. Recent developments have provided an elegant process for moving from surface topography measurements to surface BSDF. It involves the combination of the Generalised Harvey-Shack scattering theory (see §2.4.2) and the K-correlation scatter model (see §2.5.2).

The general process in this respect is as follows:

1. Measure the surface topography $z(x, y)$ and compute the PSD over the relevant spatial frequencies using Equation 2.8. For optical systems operating at short wavelengths, a combination of measurement techniques may be required to cover all relevant spatial frequencies as in [46]. These techniques include conventional surface interferometry [26], white-light surface profiling interferometry [47] and atomic force microscopy [48].

2. Fit a parametrised K-correlation model to the measured PSD. The K-correlation model is described in more detail in §2.5.2. The GHS theory of scattering is a linear systems theory and this implies (by definition) that a linear combination of K-correlation PSDs can be fitted if a single fit is inadequate.

3. Optical design software such as Zemax® may allow for direct input of the K-correlation parameters for surfaces in the lens model [64]. However, in the case of Zemax® [37], this currently means that the RR formulation for the BSDF (Equation 2.14) will be used [36]. That would also imply that only a single K-correlation model can be applied. This is generally adequate for well-polished

surfaces at moderate-long wavelengths.

4. Calculate the ACV analytically from the fitted K-correlation model(s) using Equation 2.22.

5. Compute the surface transfer function H_s for all relevant angles of incidence and scattering using Equation 2.15. Good estimates of effective and total surface roughness would be available from the surface topography/PSD measurements.

6. Calculate the ARS using Equation 2.16. If the BSDF is required, use the relation given in Equation 2.13. If the PSD spans a very large range of spatial frequencies, such as in EUV applications, the Fourier transforms used in the process have to be calculated in log space. The *FFTLog* algorithm can be used for this purpose [67].

7. The compiled ARS or BSDF can be used in optical simulations to determine the irradiance distribution in the image plane of an optical system. Of special interest in this regard is the recent work of Choi and Harvey [46, 70] in which they demonstrate computation of the imaging system PSF in analytic form in terms of convolutions of the geometrical PSF and the scaled BSDFs of the optical surfaces.

The above procedures provide a means of moving from surface profilometry (topography measurements) to a very general BSDF model that is valid for a wide range of incident and scattering angles and for smooth to moderately rough surfaces. These BSDF models can be used to compute the imaging system PSF by analytical means [46] or through raytracing with software such as Zemax®. The process described above is illustrated graphically in Figure 2.4.

The inverse problem of surface quality specification [71] is also very important. That is, given a system stray light performance metric, what roughness is allowable on the optical surfaces? However, if there is a suitable solution to a forward problem as

described in the above procedure, then solution of the inverse problem often becomes tractable merely through repeated application of the forward solution.

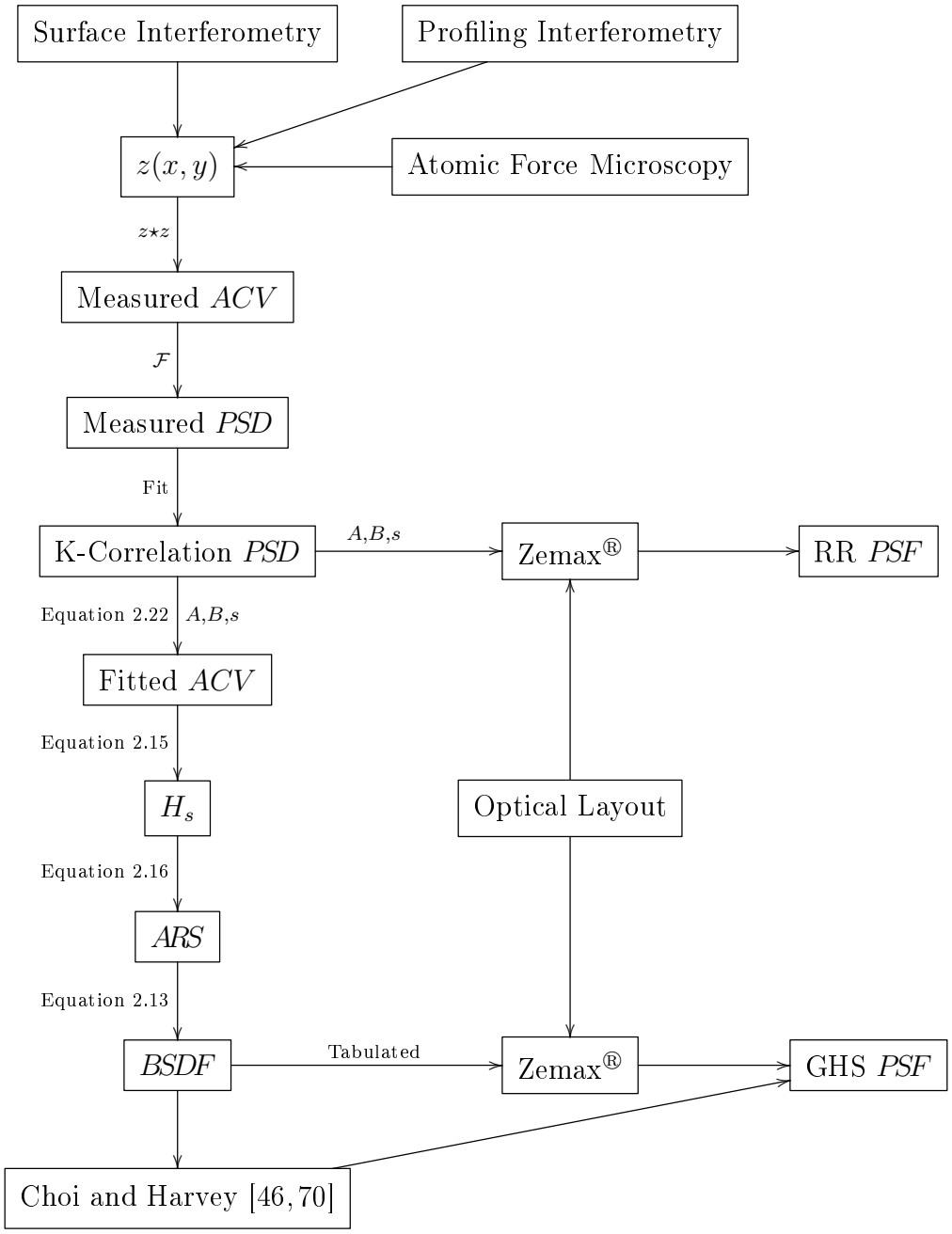

Figure 2.4: Surface Profilometry for Image Quality Modelling

Chapter 3

Measurement of Weak Scattering in the Specular Beam

3.1 Introduction

Measurement of incoherently scattered radiation very close to the coherent specular beam is challenging due to interference from the specular radiation itself [11]. Refinement of measurement techniques and scatterometer design combined with very careful geometrical definition of the source beam and sensor apertures have allowed measurements to be performed progressively closer to the specular direction [11, 21].

Many scatterometers are configured to accept a flat (plano) optical sample having a surface from which some scatter is expected [16]. The *instrument signature* is any signal generated by the scatterometer which is not attributable to scattering at the sample. The instrument signature is usually measured before insertion of the sample and subtracted from a measurement performed with the sample in place. The sample is assumed to be optically inactive except for some small incoherent scatter. This is never completely true as the sample will also introduce some fresnel reflections [22]

and optical aberrations that alter the intensity and spatial distribution of the emerging specular beam, possibly also introducing ghost beams [16].

The requirement for a plano sample is a rather limiting one, thus implying that lens and mirror surfaces with optical power cannot be directly evaluated for scatter. Instead, a surrogate, plano sample is used. This surrogate sample should be fabricated using the same process as the curved lens or mirror surfaces. Some uncertainty is therefore introduced into the process because the component used in the system is not the component that was evaluated for scattering.

Ultimately there will be an interest in the overall scattering performance of the complete optical system, possibly comprising many curved surfaces on a number of different optical substrate materials and with different thin film coatings. If the scattering characteristics of the individual surfaces are known, it is possible to model the overall performance of the system, but as noted above, there is substantial room for uncertainty. Therefore it is desirable to be in a position to measure the overall in-field stray light performance and to have a powerful means of discriminating undesirable scattered radiation from the primary specular signal. The instrument for measurement of this scattered radiation in a complete optical system will be referred to as a *system scatterometer*, as opposed to a *surface scatterometer* intended to evaluate single plano optical samples (the usual meaning of the term *scatterometer* [16]).

In this chapter, the possibility of using segmented pupil, focal plane interferometry will be explored as a means of better discriminating the in-field scattered stray light in imaging systems. This is based on the premise that there is a measurable difference between unscattered (specular, coherent) and randomly (incoherently, diffusely) scattered light. In the following development, this will be taken to mean that the specular beam exhibits interference effects while the diffusely scattered light does not.

3.2 Segmented Aperture Interferometry

There are a number of optical instruments and applications in which the pupil or aperture of an optical system is split into several segments (spatial regions) with different phase shifts applied to each segment [72]. Interference effects are then observed in the image plane of the system. An example of this technique is the Rayleigh interferometer [28]. It is commonly used to make precise measurements of the refractive indices of gases. More recently, particularly interesting examples are the dual or multiple aperture space telescope system concepts configured as Bracewell interferometers [73] for the detection of Earth-like exoplanets. This type of interferometer uses a pair of apertures on a wide baseline (20 m up to 200 m) to produce a deep interference null of extremely high viewing resolution within the image. The light from the host star is thus suppressed while light from a nearby planet can be detected and analysed. It is this same principle that is applied here to suppress the specular beam in order to measure the nearby weak incoherent scatter.

A simple type of segmented pupil, phase-shifting interferometer is one that splits the beam of an optical system in two and introduces a variable amount of phase retardation (wavefront *piston*) into one half of the beam. An optical phase retarder for this can be constructed by diamond-sawing a high quality plano-parallel glass plate in half, mounting the two pieces side-by-side in the optical beam and then rotating one half about an axis perpendicular to the cut (or about any convenient, local axis).

A simple, segmented aperture phase retarder of this type is illustrated in Figure 3.1. Rotation of one half of the retarder will alter the phase shift relative to the other half. The actual axis of rotation makes little difference, since it is the greater effective optical thickness of the rotated half that causes the phase shift.

Suppose a uniformly illuminated, plane-wave, monochromatic beam of light passes through the retarder (Figure 3.1). Assume that the slight lateral displacement of the

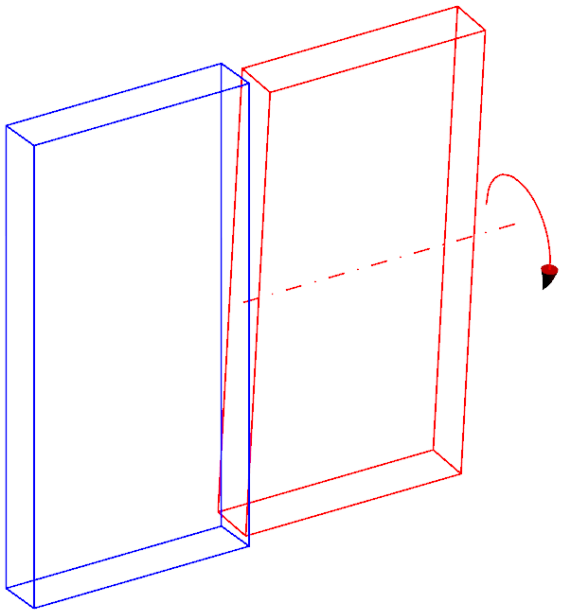

Figure 3.1: Segmented Aperture Phase Retarder

beam caused by the plate rotation is negligible. The last assumption can be realised if the system aperture stop lies beyond the retarder and the retarder is adequately overfilled by the incident beam. The phase delay in the EM field caused by the tilted plate can be written as a complex factor $e^{-i\Delta\phi}$, where $\Delta\phi$ is the phase delay expressed in radians.

The phase delay can also be written in terms of the change in optical path length through the tilted plate as a function of the tilt angle, θ_x as:

$$\Delta\phi = \frac{2\pi nt}{\lambda}\left(\frac{n}{\sqrt{n^2 - \sin^2\theta_x}} - 1\right). \tag{3.1}$$

The thickness of the plate is t and the refractive index referenced to the surrounding medium is n. A more detailed derivation of Equation 3.1 is provided in §A.1.

The far-field EM amplitude for an aperture with this phase modulation can be computed using the Fraunhofer diffraction integral. The Fraunhofer approximation can be expressed as [28]:

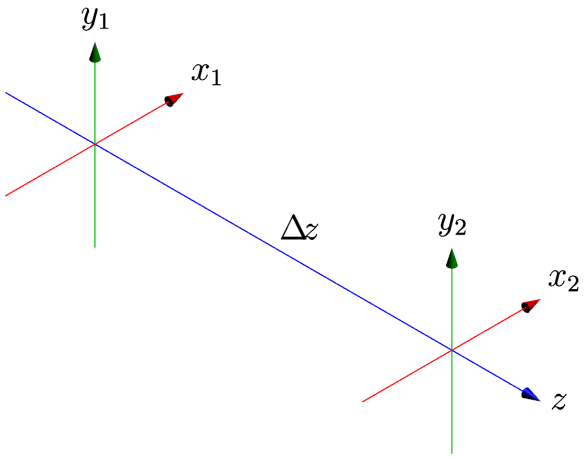

Figure 3.2: Coordinate System

$$U(x_2, y_2) = \frac{e^{ik\Delta z} e^{i\frac{k}{2\Delta z}(x_2^2+y_2^2)}}{i\lambda\Delta z} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_1, y_1) e^{-i\frac{k}{\Delta z}(x_1x_2+y_1y_2)} dx_1 dy_1, \tag{3.2}$$

where U is the scalar, time-independent field amplitude in the source plane when expressed as a function of (x_1, y_1) being the coordinates in the source plane and the scalar field amplitude in the (far-field) observation plane when expressed as a function of observation plane coordinates (x_2, y_2). The coordinate system is illustrated in Figure 3.2. It is further assumed that optical propagation occurs chiefly in the z-direction, that the propagation distance between the source and observation planes is Δz, the optical wavelength is λ and the angular wavenumber is $k = 2\pi/\lambda$.

The scalar field amplitude in the source plane $U(x_1, y_1)$, (taken to be directly after the phase retarder) can be written as the sum of two vertical apertures (strips running in the y-direction) multiplied by a single horizontal aperture (running along the x-direction) function as:

$$U(x_1, y_1) = \left[\text{rect}\left(\frac{x_1 - x_0}{D_x}\right) + \text{rect}\left(\frac{x_1 + x_0}{D_x}\right) e^{-i\Delta\phi}\right] \text{rect}\left(\frac{y_1}{D_y}\right), \tag{3.3}$$

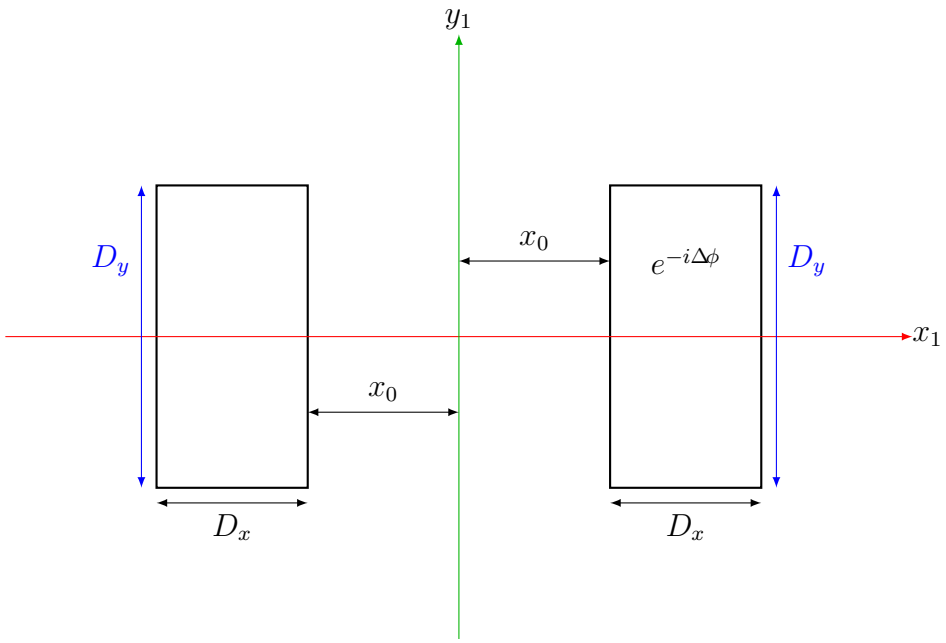

Figure 3.3: Dual Rectangular Aperture Pupil Mask

where the two vertical apertures both of width D_x are displaced by an amount x_0 on either side of the optical axis and the total vertical (y) height is D_y.

The rect function is defined as [74]:

$$\text{rect}(x) \equiv \begin{cases} 1 & |x| < \frac{1}{2} \\ \frac{1}{2} & |x| = \frac{1}{2} \\ 0 & |x| > \frac{1}{2} \end{cases}. \tag{3.4}$$

The arrangement of the two rectangular apertures (pupil segments) is shown in Figure 3.3. The constraint that $x_0 \geqq D_x/2$ is imposed so that the two apertures do not overlap.

Setting

$$u_2 = e^{ik\Delta z} e^{i\frac{k}{2\Delta z}(x_2^2+y_2^2)}, \tag{3.5}$$

the resulting Fraunhofer integral is separable as:

$$
\begin{aligned}
U(x_2,y_2) &= \frac{u_2}{i\lambda\Delta z}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_1,y_1)e^{-i\frac{k}{\Delta z}(x_1x_2+y_1y_2)}dx_1dy_1 \\
&= \frac{u_2}{i\lambda\Delta z}\int_{-\infty}^{\infty}\left[\text{rect}\left(\frac{x_1-x_0}{D_x}\right)+\text{rect}\left(\frac{x_1+x_0}{D_x}\right)e^{-i\Delta\phi}\right]e^{-i\frac{k}{\Delta z}x_1x_2}dx_1 \\
&\times \int_{-\infty}^{\infty}\text{rect}\left(\frac{y_1}{D_y}\right)e^{-i\frac{k}{\Delta z}y_1y_2}dy_1 \\
&= \frac{u_2}{i\lambda\Delta z}\left[\int_{-x_0-D_x/2}^{-x_0+D_x/2} e^{-i\frac{k}{\Delta z}x_1x_2}dx_1+e^{-i\Delta\phi}\int_{x_0-D_x/2}^{x_0+D_x/2} e^{-i\frac{k}{\Delta z}x_1x_2}dx_1\right] \\
&\times \int_{-D_y/2}^{D_y/2} e^{-i\frac{k}{\Delta z}y_1y_2}dy_1 \\
&= \frac{u_2}{i\lambda\Delta z}D_xD_y\text{sinc}\left(\frac{D_xx_2}{\lambda\Delta z}\right)\left(e^{-\frac{2\pi i}{\lambda\Delta z}x_0x_2}+e^{\frac{2\pi i}{\lambda\Delta z}x_0x_2-i\Delta\phi}\right) \\
&\times \text{sinc}\left(\frac{D_yy_2}{\lambda\Delta z}\right). \qquad (3.6)
\end{aligned}
$$

3.2.1 Focal Plane Irradiance

If the output beam from the phase retarder is focused with a perfect lens, the paraxial focal plane amplitude can be computed using the Fresnel approximation [24]. The Fresnel integral has a complex exponential quadratic term inside the integral in addition to the linear terms in the Fraunhofer integral (Equation 3.2) as:

$$
U(x_2,y_2)=\frac{u_2}{i\lambda\Delta z}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_1,y_1)e^{-i\frac{k}{\Delta z}(x_1x_2+y_1y_2)}e^{i\frac{k}{2\Delta z}(x_1^2+y_1^2)}dx_1dy_1. \qquad (3.7)
$$

However, in the paraxial approximation of a thin lens of focal length f_l, the phase change introduced in the pupil plane of the lens can also be expressed as a quadratic term of the pupil coordinates,

$$\phi_l(x_1, y_1) = -\frac{k}{2f_l}(x_1^2 + y_1^2). \tag{3.8}$$

Thus the amplitude distribution in the source plane directly after the lens becomes:

$$\begin{aligned} U(x_1, y_1) &= A(x_1, y_1)V(x_1, y_1)e^{\phi_l(x_1, y_1)} \\ &= A(x_1, y_1)V(x_1, y_1)e^{-\frac{k}{2f_l}(x_1^2 + y_1^2)}, \end{aligned}$$

where $A(x_1, y_1)$ is the incident amplitude and $V(x_1, y_1)$ is a real-valued vignetting function accounting for transmission losses of the lens. The vignetting function is introduced to cater for situations where the lens does not have equal transmittance over the whole aperture (i.e. the lens has some spatial apodization). The propagation distance to the focal plane is $\Delta z = f_l$ and the Fresnel integral becomes:

$$\begin{aligned} U(x_2, y_2) &= \frac{u_2}{i\lambda f_l} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_1, y_1)e^{-i\frac{k}{\Delta z}(x_1x_2 + y_1y_2)}e^{i\frac{k}{2\Delta z}(x_1^2 + y_1^2)}dx_1dy_1 \\ &= \frac{u_2}{i\lambda f_l} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(x_1, y_1)V(x_1, y_1)e^{-i\frac{2\pi}{\lambda f_l}(x_1x_2 + y_1y_2)}dx_1dy_1. \end{aligned} \tag{3.9}$$

The integral in Equation 3.9 can be interpreted as a Fourier transform $\mathcal{F}$ with spatial frequencies of $f_x = \frac{x_2}{\lambda f_l}$ and $f_y = \frac{y_2}{\lambda f_l}$. If the phase-retarder and dual aperture mask are placed a distance d before the lens, the phase factor outside the integral, u_2 becomes

$$u_2(x_2, y_2) = e^{i\frac{\pi}{\lambda f_l}\left(1 - \frac{d}{f_l}\right)(x_2^2 + y_2^2)}. \tag{3.10}$$

However, the vignetting factor, V, must be transformed to account for the pupil shift [75] as:

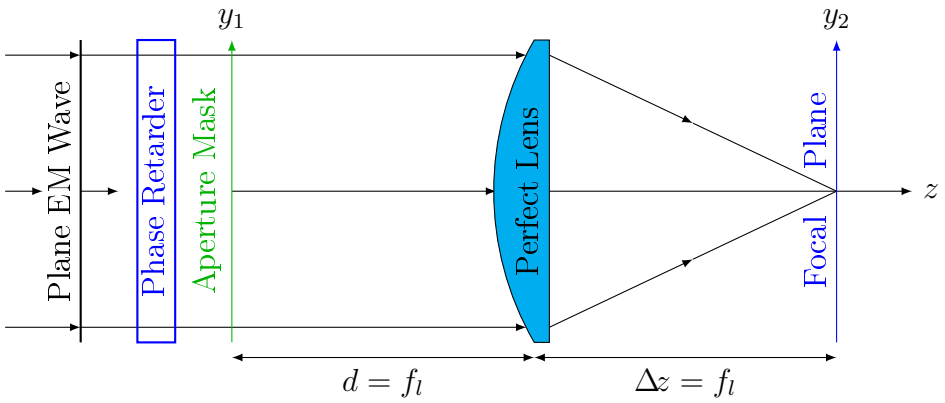

Figure 3.4: Simplified Telecentric Imaging Layout

$$U(x_2, y_2) = \frac{u_2}{i\lambda f_l}\mathcal{F}\left[A(x_1, y_1)V(x_1 + \frac{d}{f_l}x_2, y_1 + \frac{d}{f_l}y_2)\right].$$

When the phase-retarder and aperture mask are placed at the front focal point of the lens (called the *telecentric* stop position), $d = f_l$ and the phase factor $u_2 = 1$. The focal plane amplitude reduces to:

$$U(x_2, y_2) = \frac{1}{i\lambda f_l}\mathcal{F}\left[A(x_1, y_1)V(x_1 + x_2, y_1 + y_2)\right]. \tag{3.11}$$

In the cases of interest here, the lens will have a very small vignetting (wavefront amplitude spatial apodization) effect. The lens vignetting effect could be non-zero, for example, because not all rays will strike the lens surface at the same angles of incidence, and the fresnel reflection from glass is dependent on this angle [28]. Assuming the transmission/vignetting factor, V, of the lens can be neglected, and the incident amplitude, A, is set to the plane wave with segmented aperture phase retardation function given in Equation 3.3, the resulting image plane amplitude distribution follows the pattern of Equation 3.6. This simple, telecentric optical arrangement is illustrated in Figure 3.4.

Hence, under the assumed conditions of monochromatic incident plane wave, uni-

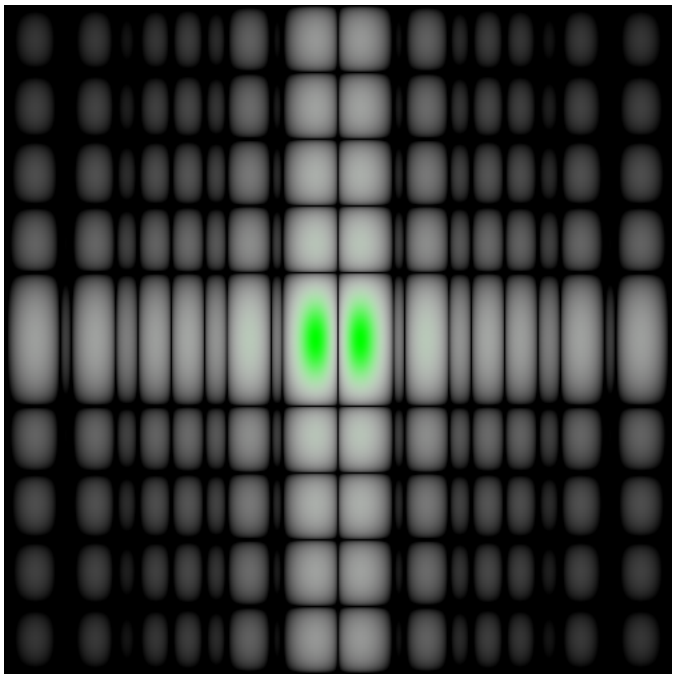

Figure 3.5: $\log E(x_2, y_2)$ for $\Delta\phi = \pi$

formly illuminated pupil as well as telecentric, paraxial and unvignetted imaging, the focal plane irradiance, $E(x_2,y_2)$, is the square modulus of the focal plane amplitude [28], $U(x_2, y_2)$ calculated according to Equation 3.11. That is:

$$\begin{aligned} E(x_2, y_2) &= \left| \frac{D_x D_y}{i\lambda f_l} \text{sinc}\left(\frac{D_x x_2}{\lambda f_l}\right) \left(e^{-\frac{2\pi i}{\lambda f_l} x_0 x_2} + e^{\frac{2\pi i}{\lambda f_l} x_0 x_2 - i\Delta\phi}\right) \text{sinc}\left(\frac{D_y y_2}{\lambda f_l}\right) \right|^2 \\ &= \left[\frac{D_x D_y}{\lambda f_l} \text{sinc}\left(\frac{D_x x_2}{\lambda f_l}\right) \text{sinc}\left(\frac{D_y y_2}{\lambda f_l}\right)\right]^2 \\ &\times \left\{2\left[\cos\left(\frac{4\pi x_0 x_2}{\lambda f_l} - \Delta\phi\right) + 1\right]\right\}. \end{aligned} \tag{3.12}$$

In the specific case of $x_2 = y_2 = 0$ and $\Delta\phi = \pi$, the axial irradiance is zero. That is, with the retarder set to one half wave, the axial irradiance drops to zero. An example image of the logarithm of $E(x_2, y_2)$ for $\Delta\phi = \pi$ is shown in Figure 3.5. The linear values are plotted in green.

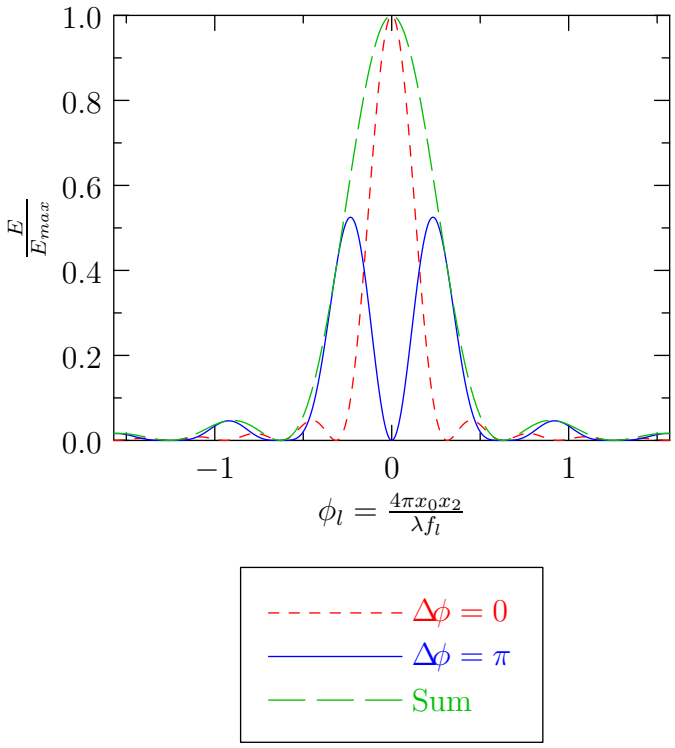

Figure 3.6: x-direction Cross Section of $E(x_2, y_2)$

A cross-section of the normalised irradiance in the x-direction through the origin is shown in Figure 3.6, where $\phi_l = \frac{4\pi x_0 x_2}{\lambda f_l}$. As might be expected, rotation of the phase retarder will give rise to a travelling interference fringe pattern. The fringes travel in the direction perpendicular to the cut through the retarder.

3.3 Source and Detector Aperture Effects

The fact that the light source and focal plane detector, in practice, will be of finite rather than infinitesimal size will have an effect on the depth of the irradiance minimum (interference null) that can be achieved. This in turn will have an important effect on the system capability to measure low levels of scattered irradiance in the interference null. In this section, these effects are explored, together with possible methods of analysis.

The result expressed in Equation 3.12 is the irradiance pattern for a uniform, monochromatic plane wave incident on the phase retarder. This is equivalent to a point source at infinity with $E(x_2, y_2)$ regarded as the system Point Spread Function (PSF). For a spatially extended source that is spatially incoherent, the linear systems approach [24] is to convolve the spatial intensity distribution of the source projected to the image plane with the system PSF. Since the intention here is to exploit the irradiance minima in the system PSF cross-section in the x-direction for the purpose of measuring scattered light, the source should compromise these minima as little as possible and should therefore have minimal spatial extent in the x-direction. A slit source with the long axis in the y-direction is therefore chosen for analysis, since this will have least impact on the depth of the irradiance null. The dimensions of the slit source, projected to the image plane will be labelled s_x and s_y. The convolution (denoted $\otimes$) of the system PSF and the projected slit aperture function, $S(x_2, y_2)$, is thus

$$\begin{aligned} E_s(x_2, y_2) &= E(x_2, y_2) \otimes S(x_2, y_2) \\ &= E(x_2, y_2) \otimes \left[\text{rect}\left(\frac{x_2}{s_x}\right) \text{rect}\left(\frac{y_2}{s_y}\right) \right]. \end{aligned}$$

Here it is assumed that the axes of the source slit are precisely aligned to the axes of the phase retarder rotation axis and mask apertures (see Figures 3.3, 3.1 and 3.4).

The convolution operation is defined [32] as,

$$f(x, y) \otimes g(x, y) \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x', y') g(x - x', y - y') dx' dy'. \tag{3.13}$$

In the case of separable functions the convolution can also be separated as:

$$
\begin{aligned}
[f_1(x)f_2(y)] \otimes [g_1(x)g_2(y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(x')f_2(y')g_1(x-x')g_2(y-y')dx'dy' \\
&= \int_{-\infty}^{\infty} f_1(x')g_1(x-x')dx' \int_{-\infty}^{\infty} f_2(y)g_2(y-y')dy' \\
&= [f_1(x) \otimes g_1(x)]\,[f_2(y) \otimes g_2(y)]\,.
\end{aligned}
$$

Likewise, if the detector has a rectangular aperture of dimensions $p_x \times p_y$ with spatially uniform response, there will be a further convolution with the detector/pixel aperture function, $P_d(x_2, y_2) = P_X(x_2)P_Y(y_2)$, which is separable for rectangular detectors. The resulting function with both detector and source convolutions can be used to determine the amount of flux falling on the detector aperture, wherever the detector is positioned in the image. The separated functions are defined as,

$$
\begin{aligned}
E_{sp}(x_2, y_2) &= E(x_2, y_2) \otimes S(x_2, y_2) \otimes P_d(x_2, y_2) \\
&= 2\left(\frac{D_x D_y}{\lambda f_l}\right)^2 [E_x(x_2) \otimes S_x(x_2) \otimes P_x(x_2)] \\
&\times\ [E_y(y_2) \otimes S_y(y_2) \otimes P_y(y_2)] \\
&= 2\left(\frac{D_x D_y}{\lambda f_l}\right)^2 E_{xsp}(x_2)E_{ysp}(y_2),
\end{aligned}
$$

where,

$$
\begin{aligned}
E_{xsp}(x_2) &= E_x(x_2) \otimes S_x(x_2) \otimes P_x(x_2) \\
&= \left\{\mathrm{sinc}^2\left(\frac{D_x x_2}{\lambda f_l}\right)\left[\cos\left(\frac{4\pi x_0 x_2}{\lambda f_l} - \Delta\phi\right) + 1\right]\right\} \\
&\otimes\ \mathrm{rect}\left(\frac{x_2}{s_x}\right) \otimes \mathrm{rect}\left(\frac{x_2}{p_x}\right)
\end{aligned}
$$

and

$$\begin{aligned} E_{ysp}(y_2) &= E_y(y_2) \otimes S_y(y_2) \otimes P_y(y_2) \\ &= \operatorname{sinc}^2\left(\frac{D_y y_2}{\lambda f_l}\right) \otimes \operatorname{rect}\left(\frac{y_2}{s_y}\right) \otimes \operatorname{rect}\left(\frac{y_2}{p_y}\right). \end{aligned}$$

These convolutions though difficult to solve analytically[1], can be computed numerically using dense sampling of x_2 and y_2 and without using Fourier methods. This result could then be compared to a numerical result computed using Fourier methods, or with Zemax® [37] which uses Fourier methods amongst others.

3.4 Bilaterally Symmetric Segmented Pupils

In this section a more general approach to analysis of segmented pupil interferometry will be explored, starting with the result that the incoherent Optical Transfer Function (OTF) can be computed as the normalised autocorrelation of the pupil function [24,25]. The generalised pupil function gives the amplitude and phase of the EM field at the exit pupil. In this case the pupil function is taken to comprise two non-overlapping segments (regions) one of which is a reflection of the other about the y-axis. Moreover, the pupil function is assumed to be symmetrical about the x-axis. The aperture mask shown in Figure 3.3 is a special case. The pupil function therefore comprises a real apodization function, $P(x, y)$, that is positive in a continuous region confined to the half plane where $x \geqslant 0$ and that is symmetrical about the x-axis, combined with the reflection of P around the y-axis multiplied by the phase factor $e^{-i\Delta\phi}$. A more general case is illustrated in Figure 3.7.

Under these conditions the generalised pupil function, $\mathbf{P}$, can be written as

[1]Solutions generally involve sine-integral, cosine-integral or exponential-integral terms.

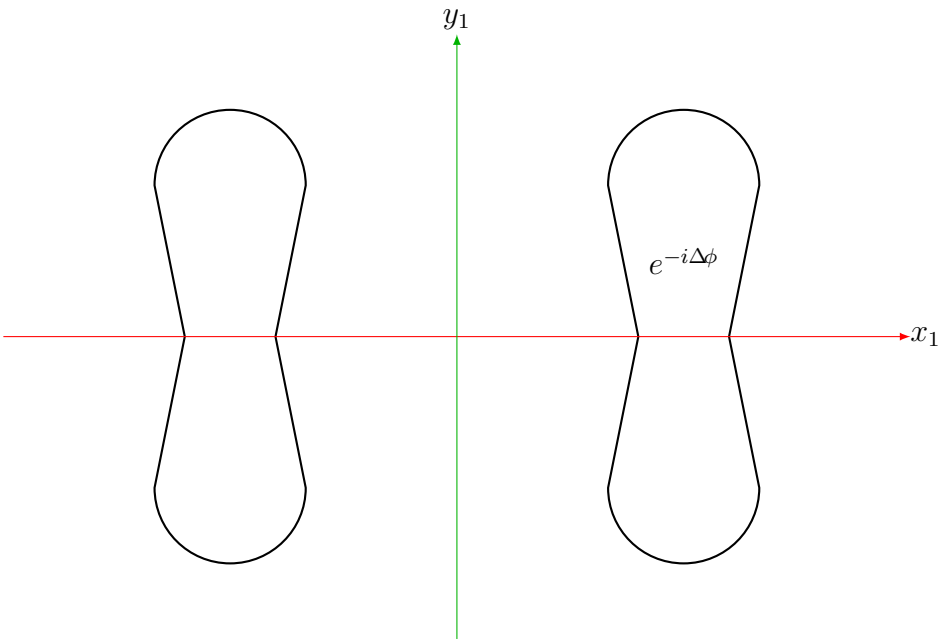

Figure 3.7: Bilaterally Symmetric Segmented Pupil Mask

$$\mathbf{P}(x,y) = P(x,y) + e^{-i\Delta\phi}P(-x,y).$$

The complex, incoherent OTF is then computed as the normalised complex autocorrelation (denoted $\star$) of the generalised pupil function. The correlation operation can be changed into a convolution operation. Doing this is advantageous because the convolution operation is commutative, associative and distributive whereas correlation is not generally commutative. The required identity [32] is

$$f(x,y) \star g(x,y) = f(-x,-y) \otimes g^*(x,y).$$

The superscript * denotes complex conjugation. The autocorrelation of $\mathbf{P}$ is then

expanded and simplified using the y-symmetry of P. Therefore:

$$\begin{aligned}
\mathbf{P}(x,y) \star \mathbf{P}(x,y) &= \left[P(x,y) + e^{-i\Delta\phi}P(-x,y)\right] \star \left[P(x,y) + e^{-i\Delta\phi}P(-x,y)\right] \\
&= P(x,y) \star P(x,y) + P(x,y) \star \left[e^{-i\Delta\phi}P(-x,y)\right] \\
&+ \left[e^{-i\Delta\phi}P(-x,y)\right] \star P(x,y) + \left[e^{-i\Delta\phi}P(-x,y)\right] \star \left[e^{-i\Delta\phi}P(-x,y)\right] \\
&= P(-x,y) \otimes P(x,y) + P(-x,y) \otimes \left[e^{i\Delta\phi}P(-x,y)\right] \\
&+ \left[e^{-i\Delta\phi}P(x,y)\right] \otimes P(x,y) + \left[e^{-i\Delta\phi}P(x,y)\right] \otimes \left[e^{i\Delta\phi}P(-x,y)\right] \\
&= P(-x,y) \otimes P(x,y) + e^{i\Delta\phi}P(-x,y) \otimes P(-x,y) \\
&+ e^{-i\Delta\phi}P(x,y) \otimes P(x,y) + P(x,y) \otimes P(-x,y) \\
&= 2P(x,y) \otimes P(-x,y) + e^{i\Delta\phi}P(-x,y) \otimes P(-x,y) \\
&+ e^{-i\Delta\phi}P(x,y) \otimes P(x,y). \qquad (3.14)
\end{aligned}$$

The incoherent PSF is proportional to the Fourier Transform of the incoherent OTF and the OTF is the autocorrelation of the pupil function (see Figure 2.2). The Fourier Transform of the autocorrelation of the generalised pupil function $\mathbf{P}$ therefore provides the PSF as:

$$\begin{aligned}
\mathcal{F}\left[\mathbf{P}(x,y) \star \mathbf{P}(x,y)\right] &\propto 2\mathcal{F}\left[P(x,y) \otimes P(-x,y)\right] + e^{i\Delta\phi}\mathcal{F}\left[P(-x,y) \otimes P(-x,y)\right] \\
&+ e^{-i\Delta\phi}\mathcal{F}\left[P(x,y) \otimes P(x,y)\right] \\
&= 2\mathcal{F}\left[P(x,y)\right]\mathcal{F}\left[P(-x,y)\right] \\
&+ e^{i\Delta\phi}\left\{\mathcal{F}\left[P(-x,y)\right]\right\}^2 + e^{-i\Delta\phi}\left\{\mathcal{F}\left[P(x,y)\right]\right\}^2. \qquad (3.15)
\end{aligned}$$

As a first check on the result in Equation 3.15, the rectangular aperture arrangement described by Equation 3.3 and illustrated in Figure 3.3 will be used and the result compared to Equation 3.12. The equivalent pupil apodization function can be written

as:

$$P(x_1, y_1) = \text{rect}\left(\frac{x_1 - x_0}{D_x}\right) \text{rect}\left(\frac{y_1}{D_y}\right).$$

The Fourier transform of this function is,

$$\mathcal{F}_\omega \left[P(x_1, y_1)\right] = D_x \text{sinc}(D_x\omega_x) e^{-i2\pi\omega_x x_0} D_y \text{sinc}(D_y\omega_y).$$

The Fourier transform of the reflection across the y-axis is

$$\mathcal{F}_\omega \left[P(-x_1, y_1)\right] = D_x \text{sinc}(D_x\omega_x) e^{i2\pi\omega_x x_0} D_y \text{sinc}(D_y\omega_y).$$

Thus

$$\begin{aligned}
\mathcal{F}_\omega \left[\mathbf{P}(x_1, y_1) \star \mathbf{P}(x_1, y_1)\right] &\propto 2\left[D_x \text{sinc}(D_x\omega_x) D_y \text{sinc}(D_y\omega_y)\right]^2 \\
&+ e^{i\Delta\phi} e^{i4\pi\omega_x x_0} \left[D_x \text{sinc}(D_x\omega_x) D_y \text{sinc}(D_y\omega_y)\right]^2 \\
&+ e^{-i\Delta\phi} e^{-i4\pi\omega_x x_0} \left[D_x \text{sinc}(D_x\omega_x) D_y \text{sinc}(D_y\omega_y)\right]^2 \\
&= \left[D_x \text{sinc}(D_x\omega_x) D_y \text{sinc}(D_y\omega_y)\right]^2 \\
&\times \left\{2 + e^{i4\pi\omega_x x_0} e^{i\Delta\phi} + e^{-i4\pi\omega_x x_0} e^{-i\Delta\phi}\right\} \\
&= 2\left[D_x \text{sinc}(D_x\omega_x) D_y \text{sinc}(D_y\omega_y)\right]^2 \\
&\times \left\{1 + \cos\left(4\pi\omega_x x_0 + \Delta\phi\right)\right\}.
\end{aligned}$$

With $\omega_x = \frac{x_2}{\lambda f_l}$ and $\omega_y = \frac{y_2}{\lambda f_l}$, this is essentially the same result to within a scaling constant and the sign of $\Delta\phi$ as that obtained in Equation 3.12. This helps to establish the validity of both Equation 3.15 and Equation 3.12.

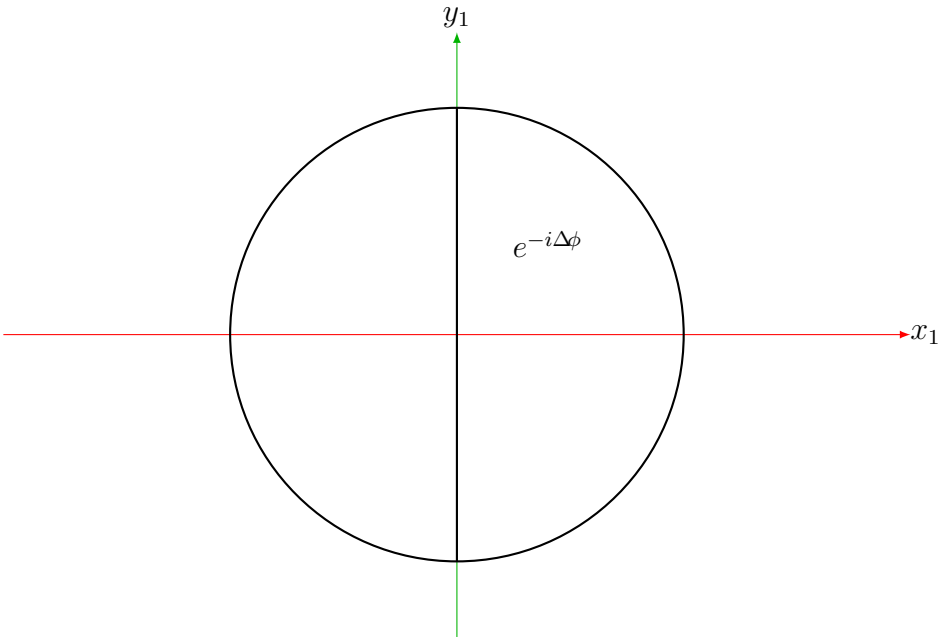

Figure 3.8: Fractional Hilbert Phase Mask

3.5 Circular Pupils

For general imaging systems, the most common pupil shape is circular. Therefore a further interesting case is that of a circular pupil, segmented into two semi-circles, one of which receives the phase shift $e^{-i\Delta\phi}$. This is more generally known as the fractional Hilbert phase mask, illustrated in Figure 3.8. A few alternative approaches were explored to reach an analytical image plane irradiance distribution expression for the fractional Hilbert pupil mask. Some of these methods involved separation of the Hilbert pupil phase mask into various combinations of circ and step functions, or expressed as infinite Fourier summations.

All of these alternative were found to have significant analytical barriers.

As an illustration, a semi-circular pupil function can be written as

$$P(x_1, y_1) = \text{circ}\left(\frac{\sqrt{x_1^2 + y_1^2}}{R}\right) \text{step}\,(x_1)\,. \tag{3.16}$$

The circ function for strictly positive radius r is expressed as follows [74]:

$$\text{circ}(r) = \begin{cases} 1 & r < 1 \\ \frac{1}{2} & r = 1 \\ 0 & r > 1 \end{cases}. \tag{3.17}$$

The Heaviside step function is defined [32] as,

$$\text{step}(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & x > 0 \end{cases}. \tag{3.18}$$

The Fourier transform of the pupil apodization function in Equation 3.16 is a convolution of the Fourier transforms of the circ and step functions. This arises from the convolution theorem, which states that the Fourier transform of a functional product is the convolution of the Fourier transforms of the factors [76]. The Fourier transform of the circ function is the jinc function or Airy disc [25], which is a scaled Bessel function of the first kind and of first order, denoted J_1, divided by the radial coordinate as:

$$\mathcal{F}_\omega \left[\text{circ}\left(\frac{\sqrt{x^2+y^2}}{R} \right) \right] = R^2 \frac{J_1\left(2\pi R\sqrt{\omega_x^2+\omega_y^2}\right)}{R\sqrt{\omega_x^2+\omega_y^2}} = R^2 \text{jinc}\left(2R\sqrt{\omega_x^2+\omega_y^2}\right),$$

where R is the pupil radius. The jinc function is defined here as:

$$\text{jinc}(\omega_x, \omega_y) = 2\frac{J_1\left(\pi\sqrt{\omega_x^2+\omega_y^2}\right)}{\sqrt{\omega_x^2+\omega_y^2}}.$$

The Fourier transform of the step function [32] is

$$\mathcal{F}_\omega\left[\text{step}(x)\right] = \frac{1}{i2\pi\omega} + \delta(\omega),$$

where δ denotes the Dirac-delta function [32].

Therefore we can write:

$$\begin{aligned}
\mathcal{F}_\omega\left[P(x,y)\right] &= \mathcal{F}_\omega\left[\text{circ}\left(\frac{\sqrt{x^2+y^2}}{R}\right)\right] \otimes \mathcal{F}_\omega\left[\text{step}(x)\right] \\
&= \mathcal{F}_\omega\left[\text{circ}\left(\frac{\sqrt{x^2+y^2}}{R}\right)\right] \otimes \left[\frac{1}{i2\pi\omega_x} + \delta(\omega_x)\right] \\
&= \mathcal{F}_\omega\left[\text{circ}\left(\frac{\sqrt{x^2+y^2}}{R}\right)\right] \otimes \left[\frac{1}{i2\pi\omega_x}\right] + \mathcal{F}_\omega\left[\text{circ}\left(\frac{\sqrt{x^2+y^2}}{R}\right)\right].
\end{aligned}$$

This Fourier transform can also be expressed as a definite integral in polar coordinates with the standard coordinate transform from Cartesian to polar coordinates [32] as:

$$\mathcal{F}_\omega[P(r,\theta)] = \int_0^R \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} r e^{-i2\pi r\omega_r \cos(\theta-\omega_\theta)} d\theta dr.$$

The generalised pupil function, $\mathbf{P}$, now including the phase factor $e^{-i\Delta\phi}$ can be written as a Fourier series in θ multiplied by the circ function so that [76]:

$$\mathbf{P}(r,\theta) = \text{circ}\left(\frac{r}{R}\right)\left\{\frac{1}{2} - \frac{2}{\pi}\sum_{n=1,3,5\ldots}^{\infty}\frac{1}{n}\sin n\theta + e^{-i\Delta\phi}\left[\frac{1}{2} + \frac{2}{\pi}\sum_{n=1,3,5\ldots}^{\infty}\frac{1}{n}\sin n\theta\right]\right\}.$$

Performing the Fourier transform of this complex pupil function is analytically challenging and thus was not taken any further.

The approaches that were attempted in reaching an analytical solution for the frac-

tional Hilbert phase mask all make the assumption that the central cut through the phase retarder (Figure 3.1) is infinitely thin. In practice, it will not be possible to implement a perfect cut in the phase retarder, so in the interest of reducing stray light, the cut in the phase retarder should be masked with a thin, rectangular obscuration. This is the equivalent of slightly shifting the position of the step function in Equation 3.16 to the right. This however only adds to the computational problems already encountered. An alternative approach to a solution for the fractional Hilbert phase mask was therefore sought and this was found in the form of the Nijboer-Zernike approach [28].

3.6 Imperfect Nulling

With perfect geometry and perfectly monochromatic and uniform illumination, the axial image irradiance is theoretically zero (interference null) for symmetric pair pupil segment arrangements with an optical phase difference of exactly π radians. In a real experiment, none of these conditions will be exactly achieved and this section will mention some of the effects of imperfections on the depth of the null.

3.6.1 Polychromaticity

No light source is perfectly monochromatic and unless the phase retarder is perfectly achromatised, the phase delay introduced by the retarder will not be exactly the same for all wavelengths. This is an inevitable consequence of the fact that all conventional optical media exhibit some dispersion [77], meaning that the refractive index varies with wavelength. In the case of a complete system, the overall spectral weighting depends on the spectral distribution of the light source, the spectral transmission of the optical train as well as the spectral response of the detector. These spectrally active effects will be rolled up into a function of wavelength called the system Spectral Response Function

(SRF), denoted $S(\lambda)$. The polychromatic PSF is then computed as an integral over wavelength of the monochromatic PSF weighted by the SRF.

In principle it is possible to achromatise the phase retarder or at least to reduce the magnitude of the dependence of retardation on wavelength. The wavelength dependence of phase retardation is expressed through the refractive index as shown in § A.1.

3.6.2 Non-Uniform Illumination or Vignetting

Non-uniform illumination of the pupil or lens vignetting effectively introduces a spatial amplitude apodization effect. This would have to be taken into consideration when analysing the effectiveness and depth of the interference null. For example, this could appear as the non-negligible lens vignetting function, V, from Equation 3.11. Evaluation of non-uniform illumination or vignetting has to be performed numerically, with software such as Zemax®. Simple functional forms of spatial apodization could however be addressed analytically.

3.6.3 Geometry Errors

Just as non-uniform illumination will result in imperfect nulling due to the reflective symmetry of the apodization function P being broken, so too will implementation errors in the effective shape or dimensions of the apertures. For example, if the edges of the apertures were not perfectly straight or orthogonal, this would have an effect on the PSF. Again, this will have to be analysed numerically.

3.6.4 Aberrations

It is not possible to construct optical systems that produce and propagate perfect optical wavefronts. Optical components in the system including lenses, windows, filters

and even the phase retarder itself will introduce non-zero optical wavefront aberrations (distortions from ideal wavefront shape). In this case the generalised pupil function must include these wavefront aberrations as a varying phase term over the aperture. Nijboer-Zernike theory is an attractive alternative for dealing with arbitrary aberrations in the propagated wavefront and the effect of such aberrations on the PSF.

3.7 Summary

A closed form solution to the focal plane EM field intensity (Equation 3.12) has been developed in this chapter for a rectangular aperture, segmented pupil (Figure 3.3), phase-shifting interferometer for a point source. In practice, convolutions of this solution would be required to take finite source and detector sizes into account. Closed form solution of these convolutions was not pursued here and a numerical approach would be the first choice in further work on this problem.

Analysis of the fractional Hilbert mask is further complicated by the computational barriers encountered. The above alternatives lead to a search for a completely different approach in solving this problem. This was found in the form of the Nijboer-Zernike theory which enables computation of the focal plane field through expansion of the pupil function in terms of Zernike polynomials [28]. Further, in this investigation, the extended Nijboer-Zernike approach (ENZ) that was encountered [78] expands the scope of problems that can be addressed using Zernike-basis expansion.

The analysis of the fractional Hilbert mask in the context of ENZ theory resulted in a journal paper submission to Current Applied Physics entitled "Zernike-basis expansion of the fractional and radial Hilbert phase masks". The paper is reproduced in its entirety in Chapter 6.

Chapter 4

Precision Imaging Systems

4.1 Introduction

Precision optical imaging systems are used in a very broad range of applications. Some of the most demanding imaging applications are those in microscopy, astronomy, space optics and lithography. Design and construction of objective lens assemblies for ultraviolet lithography [15, 18–20] is a key technology in the manufacture of state-of-the-art electronic Integrated Circuits (IC). To illustrate the scale and complexity of these systems, a lithography objective lens system patent *circa* 1998 [79] is shown in cross-section in Figure 4.1. The completed opto-mechanical assembly of the Zeiss® Starlith 1000 lithography lens is shown in Figure 4.2, which operates at a wavelength of $\lambda = 248$ nm [80].

The lithography lens objective system illustrated in Figure 4.1 comprises 59 refractive lens surfaces. Surface quality and cleanliness must be controlled to very high standards in order to minimise the effects of stray light.

A variety of phenomena can degrade the contrast or quality of the image produced by a precision optical system. To illustrate the extensive range of these effects, the

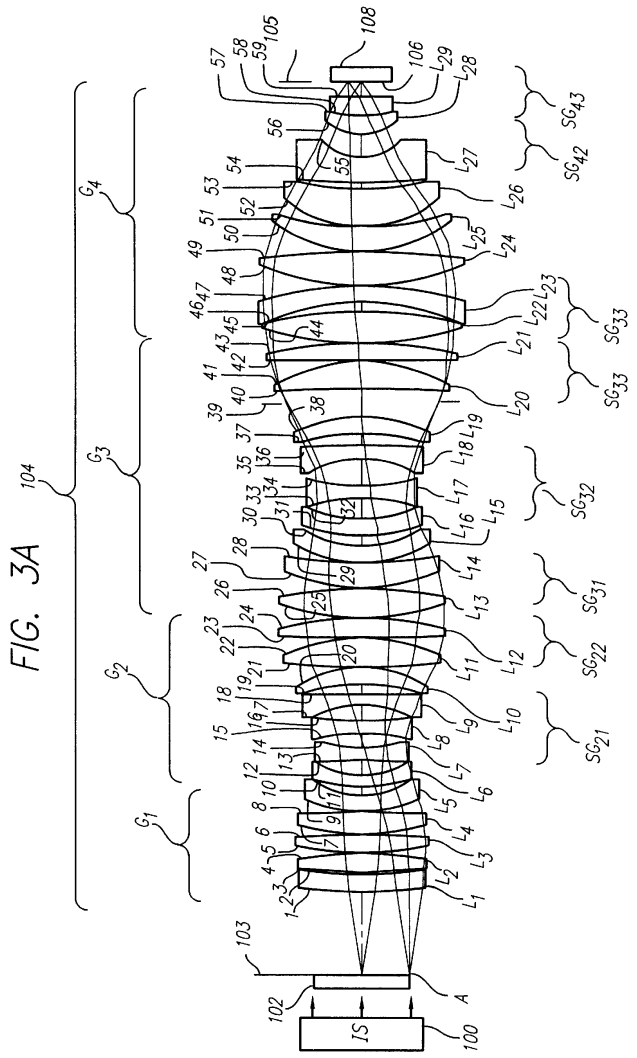

Figure 4.1: Lithography Lens Optical Design Cross-Section, USP 5,969,803

Figure 4.2: Zeiss® Starlith 1000 Lithography Lens Opto-Mechanical Assembly, Courtesy Electronics Production and Test [80]

Figure 4.3: Moon Image with Low Stray Light (left) and Simulated Stray Light (right)

following sections provide an introductory discussion, with a broad distinction into design-related, manufacture-related and use-related effects.

While there are a great number of potential degrading effects, the unifying theme is that of absorption or directional scatter of light. While absorption only reduces the available light for the application, undesirable scattered light reaching the image plane will reduce image quality. The general term for such undesirable scattered light is *stray light* [6]. Figure 4.3 is a simulation of the effect that stray light could have on a high quality image.

4.2 Optical Design and Opto-mechanical Design

The optical design of an imaging system is the primary determinant of the limiting image quality that a system can deliver [81]. In the regime of geometrical optics, system performance is analysed by performing geometrical tracing of rays through the optical system according to the laws of refraction (Snell's law) and reflection [22]. Any deviations from perfect intersection by the rays at a geometrical point in the image

plane are referred to as geometrical ray *aberrations* [82]. The *chromatic* aberrations are those related to the fact that optical glasses have a different refractive index for different wavelengths of light. This is also referred to as optical *dispersion* [77, 83]. Generally, glass components also exhibit a change in dimensions as well as refractive index with temperature change [84]. Hence, glass must be selected in the design process to compensate both for dispersion and other physical and chemical properties of the glass [85, 86].

Opto-mechanical design is the process of designing a structure to provide static or dynamic (e.g. zoom lens) positional location and support to the active optical components (lenses, mirrors, prisms, filters etc.) of an optical system [87]. The manner of support has an influence on such things as the self-weight distortion of an optical component under the effect of gravity. Components can shift as a result of temperature fluctuations as the mechanical components expand and contract. Stress can be induced through the mechanical pinching of optical components leading to birefringence [88], which gives rise to double images [89].

Opto-mechanical components provide limiting apertures and other support structures such as spiders and mirror cells that diffract and absorb light passing through the system. Light can scatter from mechanical components and reach the image plane, reducing contrast. This problem is generally controlled using carefully formulated black paints or coatings [35, 90] and by including baffles which block the light scattered from mechanical parts [6].

All glass materials exhibit a residual surface reflection known as the *fresnel* reflection [22]. These reflections can give rise to ghost images which interfere with primary image utility. Ghost images are controlled by manipulating the geometrical layout of optical components, as well as the application of thin optical films (anti-reflection coatings). These thin film coatings can have other influences, such as increasing random scatter [91]

and distortion of the component through increased stress.

With due consideration to the image quality demanded of a system as well as the operating environment in which the system must maintain performance, the combined optical and opto-mechanical designs must be configured to within manufacturing tolerance bands to meet these requirements [87].

4.3 Optical Manufacture

Raw materials used in the production of precision optical systems are produced by industrial processes which cannot yield materials that are perfect in every respect. For example, optical glasses cannot be reproduced to have a perfectly precise refractive index or with perfect spatial homogeneity [77]. Spatial fluctuations in refractive index can occur on scales of a fraction of a millimetre (*striae*) to gradients in refractive index across a large glass sample [92, 93]. Glass can contain chemical impurities and inclusions (bubbles, crystals or solid foreign matter) which absorb or scatter light [94]. While glass for precision optical systems is generally subject to the process of *annealing* [77] to reduce internal stresses, there will always be some residual internal stress after annealing [88].

Manufacture of precision optical components from the raw materials is executed using a very broad range of industrial processes from traditional grind and pitch-lap polishing [66] to state-of-the-art techniques such as Magneto-Rheological Finishing (MRF [95]). The physical and chemical characteristics of the raw material in conjunction with the selected manufacturing processes results in a component which deviates from the perfect geometry and optical properties of the designed component. Residual surface roughness (imperfect polish) and flaws such as scratches and digs as well as sub-surface damage [96] such as micro-cracks will also cause undesirable scattering.

The residual surface shape errors are often classified as *form* (or *figure*), being the low spatial frequencies, *ripple* (mid-spatial frequencies) and *finish* (roughness, high spatial frequencies) [97].

System integration involves the construction of the optical sub-assemblies and the system from the components. The methods, controls and processes used in system assembly can affect the quality of the resulting system [87].

Control of contamination from dust or other undesirable substances such as deposited airborne Volatile Organic Compounds (VOC) can be a very important aspect of the manufacturing process [98, 99], especially in aerospace applications [100].

These manufacturing errors and deficiencies, both in optical and opto-mechanical components and assemblies will generally result in degraded image quality.

4.4 Operational Environment

Transition of the optical system into the operational environment, as well as the effects of the operational environment itself can cause degradation in imaging performance. An example of a severe transition environment is the launch of a space optical system [101], with extreme levels of acceleration, shock, vibration and temperature fluctuation.

Once an optical system is operating in the intended environment there are various mechanisms by which imaging performance can be degraded [102]. While optical glasses are generally very stable materials, they are susceptible to damage and ageing effects [77]. These effects would include chipping, scratching, devitrification, solarisation, hard radiation damage, surface etching and water damage. The optical thin film coatings applied to lens and mirror surfaces may delaminate or disintegrate due to one of the above processes. Environmental contaminants, such as fingerprints and dust can cause etching of coatings and underlying glass surfaces.

Solarisation is generally reversible damage to the spectral transmission of a transparent material resulting from intense exposure to EM radiation (soft X-rays through to ultraviolet wavelengths). Glasses used in space applications are usually subject to higher levels of ultraviolet and shorter wavelength radiation in general.

If glass is subject to high fluxes of EM radiation such as from a laser, specific types of surface and bulk damage become a risk [103]. Existing damage as well as contamination tend to increase the risk of further damage.

Hard radiation (e.g. protons, gamma rays) such as may be encountered in a space environment can cause damage to glass types not formulated to be resistant to such exposure [104, 105].

Temperature fluctuations not only cause variations in refractive index, but non-uniform temperature distributions can cause component deformation and additional stress-induced birefringence [88].

While the above effects relate to optical scatter from features occurring at various physical size scales, it is important to note that there is also background scatter caused at the fundamental physical granularity (atoms, molecules) of the materials comprising the optical system. Elastic atomic and molecular scatter is also referred to as Rayleigh scattering [5, 45, 106], which is relatively isotropic (the same over all scatter angles). In particular, the gases between optical components as well as the glass materials will exhibit both elastic and inelastic molecular scatter within the bulk of the material. While Rayleigh scattering is usually at a substantially lower level of intensity than the larger scale mechanisms discussed above, inelastic or non-linear scatter such as Raman [45, 107, 108] scatter is usually of substantially lower intensity again than Rayleigh scatter.

Little can be done to reduce molecular scatter. The optical system can be evacuated to eliminate scattering from gases (normally the default situation in space optical

systems). This has to be taken into account during the design phase. In the case of EUV lithographic systems, evacuation of the system is mandatory because air strongly absorbs EM radiation at the operating wavelength of $\lambda = 13.5$ nm.

4.5 Summary

Optical imaging systems are used in a large variety of applications, some of which place extreme demands on image quality. DUV/EUV lithography systems and space telescopes are prominent examples. Optical systems in this class are vulnerable at every stage of design, construction and deployment to factors which can degrade the quality and utility of the images they produce. Some of these factors have been explored in this chapter and clearly highlight the need to test and guarantee the quality of materials, components, sub-assemblies and systems at every possible opportunity. This applies equally to the procedures and processes used to create these systems as well as the manufacturing and operational environments. Directional scattering of light plays a central role and is the unifying principle across all of these considerations. Theoretical understanding, modelling and measurement of optical scattering at material, component and systems level are all mandatory in order to realise the full potential of these systems.

Chapter 5

Conclusion

This work undertook to review the applications, measurement and modelling of the directional scatter of light, with an emphasis on precision optical imaging systems. Diffraction, reflection and refraction were identified as specific types of coherent scattering along with random scatter due to optical surface defects, residual roughness, sub-surface damage, coating structure and contamination amongst others.

The problem of stray light in high performance imaging systems was investigated and it was found that by using a combination of the Generalised Harvey-Shack scatter theory with the K-correlation scatter model, it is possible to calculate a surface BSDF from profile measurements that is valid over a broad range of optical wavelengths, incident and scattering angles as well as roughness levels. These models are suitable for use in the design and analysis of stray light effects in precision imaging systems.

It was further determined that scatterometer design for measurement of weak scatter as well as construction and operation of such scatterometers is inherently challenging. A concept using segmented pupil interferometry (for example the fractional Hilbert phase mask) has been proposed to cause a dark interference fringe to traverse the specular beam, facilitating measurement of weak incoherent scatter near or within the specular

beam.

A large variety of possible deleterious effects related to design, manufacture and use of imaging systems was observed. These effects range from design flaws to optical surface contamination. The unifying theme for all of these effects is the coherent or random directional scattering of light.

Furthermore, it was found that techniques based on the ENZ approach offer considerable promise for the purpose of analysing measurements of the system PSF in order to quantify (retrieve) the various aberrations present in the imaging system.

A fractional Hilbert phase mask has been proposed and analysed as a means to facilitate measurement of weak scatter within or near the specular beam. The mask can also serve to enhance the ENZ aberration retrieval process by means of improving PSF measurement diversity. These considerations lead to a paper entitled "Zernike-basis expansion of the fractional and radial Hilbert phase masks", which is provided in Chapter 6.

Bibliography

[1] J. J. Koenderink, A. J. van Doorn, and M. Stavridi, "Bidirectional reflection distribution function expressed in terms of surface scattering modes," in *Computer Vision ECCV '96* (B. Buxton and R. Cipolla, eds.), vol. 1065 of *Lecture Notes in Computer Science*, pp. 28–39, Springer, Berlin, Heidelberg, 1996. http://dx.doi.org/10.1007/3-540-61123-1_125.

[2] R. Montes and C. Ureña, "An overview of BRDF models," 2010. http://digibug.ugr.es/bitstream/10481/19751/1/rmontes_LSI-2012-001TR.pdf, Accessed 2015-02-26.

[3] R. Pacanowski, O. Celis, C. Schlick, X. Granier, P. Poulin, and A. Cuyt, "Rational BRDF," *Visualization and Computer Graphics, IEEE Transactions on*, vol. PP, no. 99, p. 1, 2012.

[4] Y. Govaerts and M. Verstraete, "Raytran: a monte carlo ray-tracing model to compute light scattering in three-dimensional heterogeneous media," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 36, no. 2, pp. 493–505, 1998. http://dx.doi.org/10.1109/36.662732.

[5] G. E. Thomas and K. Stamnes, *Radiative transfer in the atmosphere and ocean*. Cambridge University Press, Cambridge, UK, 2002.

[6] R. P. Breault, *Handbook of Optics*, vol. 1, ch. 38 : Control of Stray Light, pp. 38.1–38.35. McGraw-Hill, New York, USA, 2nd ed., 1995.

[7] E. C. Fest, *Stray Light Analysis and Control*. SPIE Press, Bellingham, Washington, USA, 1st ed., 2013. http://dx.doi.org/10.1117/3.1000980.

[8] S. Liang, A. H. Strahler, M. J. Barnsley, C. C. Borel, S. A. W. Gerstl, D. J. Diner, A. J. Prata, and C. L. Walthall, "Multiangle remote sensing: Past, present and future," *Remote Sensing Reviews*, vol. 18, no. 2-4, pp. 83–102, 2000.

[9] R. Latifovic, J. Cihlar, and J. Chen, "A comparison of BRDF models for the normalization of satellite optical data to a standard sun-target-sensor geometry," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 41, pp. 1889 – 1898, aug. 2003.

[10] S. Liang, *Quantitative remote sensing of land surfaces*, vol. 30. John Wiley & Sons, Hoboken, NJ, USA, 2005.

[11] J. C. Stover, *Optical Scattering: Measurement and Analysis*. Bellingham, WA: SPIE Press, Bellingham, Washington, USA, second ed., 1995. http://dx.doi.org/10.1117/3.203079.

[12] C. F. Bohren and D. R. Huffman, *Absorption and scattering of light by small particles*. Wiley, Weinheim, Germany, 2nd ed., 1998.

[13] F. E. Nicodemus, J. C. Richmond, J. Hsia, I. Ginsberg, and T. Limperis, "Geometrical considerations and nomenclature for reflectance," tech. rep., NIST, October 1977. http://www.researchgate.net/publication/215506168_Geometrical_considerations_and_Nomenclature_for_Reflectance, Accessed 2015-02-26.

[14] L. Feinberg, L. Cohen, B. Dean, W. Hayden, J. Howard, and R. Keski-Kuha, "Space telescope design considerations," *Optical Engineering*, vol. 51, no. 1, p. 011006, 2012.

[15] T. Matsuyama, Y. Ohmura, and D. M. Williamson, "The lithographic lens: its history and evolution," in *Proc. SPIE*, vol. 6154, pp. 615403–615403–14, 2006. http://dx.doi.org/10.1117/12.656163.

[16] J. C. . Stover, *Handbook of Optics*, vol. 2, ch. 26: Scatterometers. McGraw-Hill, New York, USA, 2nd ed., 1995.

[17] V. B. Podobedov, M. E. Nadal, and C. C. Miller, "Improving the performance of the NIST five axis goniospectrometer for measurements of bidirectional reflectance distribution function," vol. 8065, p. 80651I, SPIE, 2011. http://dx.doi.org/10.1117/12.882865.

[18] Y.-C. Kim, P. D. Bisschop, and G. Vandenberghe, "Evaluation of stray light and quantitative analysis of its impact on lithography," *Journal of Microlithography, Microfabrication, and Microsystems*, vol. 4, no. 4, p. 043002, 2005.

[19] B. Wu and A. Kumar, "Extreme ultraviolet lithography: A review," *Journal of Vacuum Science & Technology B*, vol. 25, no. 6, pp. 1743–1761, 2007. http://dx.doi.org/10.1116/1.2794048.

[20] V. Bakshi, *EUV Lithography*, vol. 178. SPIE Press, Bellingham, Washington, USA, 2009. http://dx.doi.org/10.1117/3.769214.

[21] K. Tayabaly, J. C. Stover, R. E. Parks, M. Dubin, and J. H. Burge, "Use of the surface PSD and incident angle adjustments to investigate near specular scatter from smooth surfaces," in *Proc. SPIE*, vol. 8838, pp. 883805–883805–10, 2013. http://dx.doi.org/10.1117/12.2024612.

[22] F. Jenkins and H. White, *Fundamentals of Optics*. International student edition, McGraw-Hill, New York, USA, 1976.

[23] R. G. Driggers, P. Cox, and T. Edwards, *Introduction to Infrared and Electro-Optical Systems*. Artech House, Norwood, MA, USA, 1999.

[24] J. W. Goodman, *Introduction to Fourier Optics*. Physical and Quantum Electronics, McGraw-Hill, New York, USA, 1968.

[25] A. Papoulis, *Systems and Transforms with Applications in Optics*. McGraw-Hill Series in System Science, McGraw-Hill, New York, USA, 1968.

[26] D. Malacara, *Optical Shop Testing*. Wiley Series in Pure and Applied Optics, John Wiley & Sons, Hoboken, NJ, USA, 2nd ed., 1992.

[27] D. Malacara, M. Servín, and Z. Malacara, *Interferogram Analysis for Optical Testing, Second Edition*. Optical Engineering - Marcel Dekker, Taylor & Francis Group, London, UK, 2005.

[28] M. Born and E. Wolf, *Principles of Optics*. Cambridge University Press, Cambridge, UK, 7th ed., 1999.

[29] X. Colonna De Lega and P. de Groot, "Lateral resolution and instrument transfer function as criteria for selecting surface metrology instruments," in *Imaging and Applied Optics Technical Papers*, p. OTu1D.4, Optical Society of America, 2012. http://dx.doi.org/10.1364/OFT.2012.OTu1D.4.

[30] G. Schaepman-Strub, M. Schaepman, T. Painter, S. Dangel, and J. Martonchik, "Reflectance quantities in optical remote sensing–definitions and case studies," *Remote Sensing of Environment*, vol. 103, pp. 27–42, July 2006.

[31] W. Osten and N. Reingand, *Optical Imaging and Metrology: Advanced Technologies*. John Wiley & Sons, Hoboken, NJ, USA, 2012.

[32] E. Weisstein, *CRC Concise Encyclopedia of Mathematics, Second Edition*. Taylor & Francis Group, London, UK, 2002.

[33] R. Lyons, *Understanding Digital Signal Processing*. Prentice Hall, Upper Saddle River, NJ, USA, 2010.

[34] S. K. Kimothi, *The Uncertainty of Measurements: Physical and Chemical Metrology: Impact and Analysis*. ASQ Quality Press, Milwaukee, WI, USA, 2002.

[35] S. M. Pompea and R. P. Breault, *Handbook of Optics*, vol. 1, ch. 37 : Black Surfaces for Optical Systems, pp. 37.1–37.70. McGraw-Hill, New York, USA, 2nd ed., 1995.

[36] M. G. Dittman, "K-correlation power spectral density and surface scatter model," in *Proc. SPIE* (O. M. Uy, S. A. Straka, J. C. Fleming, and M. G. Dittman, eds.), vol. 6291, p. 62910R, SPIE, 2006. http://dx.doi.org/10.1117/12.678320.

[37] RadiantZemax, *ZEMAX Optical Design Program User's Manual*, april 14, 2010 ed., 2010. http://www.radiantzemax.com/zemax, Accessed 2015-02-26.

[38] J. E. Harvey, S. Schröder, N. Choi, and A. Duparré, "Total integrated scatter from surfaces with arbitrary roughness, correlation widths, and incident angles," *Optical Engineering*, vol. 51, no. 1, p. 013402, 2012.

[39] L. Mazule, S. Liukaityte, R. C. Eckardt, A. Melninkaitis, O. Balachninaite, and V. Sirutkaitis, "A system for measuring surface roughness by total integrated scattering," *Journal of Physics D: Applied Physics*, vol. 44, no. 50, p. 505103, 2011. http://dx.doi.org/10.1088/0022-3727/44/50/505103.

[40] SEMI, "Test Method for Measuring the Reflective Total Integrated Scatter," 2005. http://ams.semi.org/ebusiness/standards/SEMIStandardDetail.aspx?ProductID=211&DownloadID=1513, Accessed 2015-01-29.

[41] P. L. Gourley, L. R. Dawson, T. M. Brennan, B. E. Hammons, J. C. Stover, C. F. Schaus, and S. Sun, "Optical scatter in epitaxial semiconductor multilayers," *Applied Physics Letters*, vol. 58, pp. 1360 –1362, apr 1991.

[42] Z. Zhenrong, Z. Jing, and G. J. Peifu, "Roughness characterization of well-polished surfaces by measurements of light scattering distribution," *Optica Applicata*, vol. XL, no. 4, pp. 811–818, 2010.

[43] J. E. Harvey, S. Shröder, N. Choi, and A. Duparré, "Determining parametric TIS behavior from optical fabrication metrology data," in *Proc. SPIE* (J. H. Burge, O. W. Fahnle, and R. Williamson, eds.), vol. 8126, p. 81260X, SPIE, 2011.

[44] S. Schröder, A. Duparré, L. Coriand, A. Tünnermann, D. H. Penalver, and J. E. Harvey, "Modeling of light scattering in different regimes of surface roughness," *Opt. Express*, vol. 19, pp. 9820–9835, May 2011.

[45] G. Fuxi, *Optical and spectroscopic properties of glass*. Springer, Berlin, Germany, 1992.

[46] N. Choi and J. E. Harvey, "Image degradation due to surface scatter in the presence of aberrations," *Appl. Opt.*, vol. 51, pp. 535–546, Feb 2012. http://dx.doi.org/10.1364/AO.51.000535.

[47] J. Schmit, K. Creath, and J. C. Wyant, *Optical Shop Testing*, ch. 15. Surface Profilers, Multiple Wavelength, and White Light Interferometry, pp. 667–755. John Wiley & Sons, Hoboken, NJ, USA, 2006. http://dx.doi.org/10.1002/9780470135976.ch15.

[48] C. Ruppe and A. Duparré, "Roughness analysis of optical films and substrates by atomic force microscopy," *Thin Solid Films*, vol. 288, no. 1-2, pp. 8–13, 1996. http://dx.doi.org/10.1016/S0040-6090(96)08807-4.

[49] C. C. Asmail, C. L. Cromer, J. Proctor, and J. J. Hsia, "Instrumentation at the National Institue of Standards and Technology for bidirectional reflectance distribution function (BRDF) measurements," vol. 2260, pp. 52–61, SPIE, 1994. http://dx.doi.org/10.1117/12.189203.

[50] G. Obein, R. Bousquet, and M. E. Nadal, "New NIST reference goniospectrometer," in *Proc. SPIE*, vol. 5880, pp. 58800T–10, 2005. http://dx.doi.org/10.1117/12.621516.

[51] H. W. Yoon, D. W. Allen, G. P. Eppeldauer, and B. K. Tsai, "The extension of the NIST BRDF scale from 1100 nm to 2500 nm," in *Proc. SPIE* (J. J. Butler, X. Xiong, and X. Gu, eds.), vol. 7452, p. 745204, SPIE, 2009. http://dx.doi.org/10.1117/12.827293.

[52] H. J. Patrick, C. J. Zarobila, and T. A. Germer, "The NIST Robotic Optical Scatter Instrument (ROSI) and its application to BRDF measurements of diffuse reflectance standards for remote sensing," in *Proc. SPIE*, vol. 8866, pp. 886615–886615–12, 2013. http://dx.doi.org/10.1117/12.2023095.

[53] N. Choi and J. E. Harvey, "Comparison of the domain of validity of several approximate surface scatter theories," in *Proc. SPIE*, vol. 8495, pp. 849504–849504–12, 2012. http://dx.doi.org/10.1117/12.930558.

[54] J. E. Harvey, A. Krywonos, and C. L. Vernold, "Modified Beckmann-Kirchhoff scattering model for rough surfaces with large incident and scattering angles," *Optical Engineering*, vol. 46, no. 7, p. 078002, 2007.

[55] S. Rice, "Reflection of electromagnetic waves from slightly rough surfaces," *Commun. Pure Appl. Math.*, vol. 4, no. 2-3, pp. 351–378, 1951.

[56] C. L. Vernold and J. E. Harvey, "Comparison of Harvey-Shack scatter theory with experimental measurements," in *Proc. SPIE* (Z.-H. Gu and A. A. Maradudin, eds.), vol. 3141, pp. 128–138, SPIE, 1997.

[57] J. E. Harvey, A. Krywonos, and J. C. Stover, "Unified scatter model for rough surfaces at large incident and scatter angles," in *Proc. SPIE*, vol. 6672, pp. 66720C–66720C–8, 2007. http://dx.doi.org/10.1117/12.739139.

[58] N. Choi and J. E. Harvey, "Numerical validation of the generalized Harvey-Shack surface scatter theory," *Optical Engineering*, vol. 52, no. 11, pp. 115103–115103, 2013. http://dx.doi.org/10.1117/1.OE.52.11.115103.

[59] A. Krywonos, J. E. Harvey, and N. Choi, "Linear systems formulation of scattering theory for rough surfaces with arbitrary incident and scattering angles," *J. Opt. Soc. Am. A*, vol. 28, pp. 1121–1138, Jun 2011.

[60] J. E. Harvey, *Light-Scattering Characteristics of Optical Surfaces.* PhD thesis, 1976. University of Arizona, USA.

[61] E. L. Church and P. Z. Takacs, "Optimal estimation of finish parameters," in *Proc. SPIE* (J. C. Stover, ed.), vol. 1530, pp. 71–85, SPIE, 1991.

[62] S. Gangadhara, "What Scattering Models are Available in Zemax ?," 2010. http://www.zemax.com/support/knowledgebase/what-scattering-models-are-available-in-zemax, Accessed 2015-01-31.

[63] E. L. Church and P. Z. Takacs, "Light scattering from non-gaussian surfaces," in *Proc. SPIE* (J. C. Stover, ed.), vol. 2541, pp. 91–107, SPIE, 1995.

[64] S. Gangadhara, "How to Model Surface Scattering via the K-Correlation Distribution," March 2010. http://www.zemax.com/support/knowledgebase/how-to-model-surface-scattering-via-the-k-correlat, Accessed 2015-01-31.

[65] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*, vol. 55 of *Applied Mathematics Series*. NIST, Gaithersburg, MD, USA, 10th ed., 1964. http://www.cs.bham.ac.uk/~aps/research/projects/as/book.php, Accessed 2015-02-26.

[66] D. F. Horne, *Optical production technology*, vol. 1. A. Hilger, London, UK, 1972.

[67] J. E. Harvey, N. Choi, A. Krywonos, and J. G. Marcen, "Calculating BRDFs from surface PSDs for moderately rough optical surfaces," in *Proc. SPIE* (J. H. Burge, O. W. Fahnle, and R. Williamson, eds.), vol. 7426, p. 74260I, SPIE, 2009.

[68] E. L. Church and P. Z. Takacs, "Specification of the surface figure and finish of optical elements in terms of system performance," in *Proc. SPIE* (L. R. Baker, ed.), vol. 1781, pp. 118–130, SPIE, 1993.

[69] J. E. Harvey, N. Choi, A. Krywonos, and G. Peterson, "Predicting image degradation from optical surface metrology data," in *Optical Fabrication and Testing*, p. OMB2, Optical Society of America, 2010.

[70] N. Choi and J. E. Harvey, "Linear systems formulation of image analysis in the presence of both aberrations and surface scatter," in *Proc. SPIE* (R. B. Johnson, V. N. Mahajan, and S. Thibault, eds.), vol. 8128, p. 81280B, SPIE, 2011.

[71] J. Stover and S. Han, "The art of specifying surface quality," in *Proc. SPIE*, vol. 9276, pp. 92760M–92760M–5, 2014. http://dx.doi.org/10.1117/12.2071006.

[72] J. Scrimgeour and J. E. Curtis, "Aberration correction in wide-field fluorescence microscopy by segmented-pupil image interferometry," *Opt. Express*, vol. 20, pp. 14534–14541, Jun 2012. http://dx.doi.org/10.1364/OE.20.014534.

[73] D. Defrere, O. Absil, C. Hanot, and M. Fridlund, "Potential of space-based infrared Bracewell interferometers for planet detection," vol. 6693, p. 66931N, SPIE, 2007.

[74] D. Voelz, *Computational Fourier Optics : A MATLAB Tutorial*. SPIE Press, Bellingham, Washington, USA, 2010.

[75] J. D. Schmidt, *Numerical Simulation of Optical Wave Propagation*. SPIE Press, Bellingham, Washington, USA, 2010.

[76] K. Howell, *Principles of Fourier Analysis*. Studies in Advanced Mathematics, CRC Press, Taylor & Francis Group, London, UK, 2010.

[77] P. Hartmann, *Optical Glass*, vol. PM249. SPIE Press, Bellingham, Washington, USA, 2014.

[78] A. J. E. M. Janssen, "Extended Nijboer–Zernike approach for the computation of optical point-spread functions," *J. Opt. Soc. Am. A*, vol. 19, pp. 849–857, May 2002.

[79] R. I. Mercado, "Large NA projection lens for excimer laser lithographic systems," Oct. 19 1999. US Patent 5,969,803 https://www.google.com/patents/US5969803, Accessed 2015-02-26.

[80] Zeiss International, "Lithography at 248 nm (KrF)," 2015. http://www.zeiss.com/semiconductor-manufacturing-technology/

en_de/products-solutions/lithography-optics/lithography-at-248-nanometers--krf-.html, Accessed 2015-02-26.

[81] W. J. Smith, *Modern Optical Engineering*, vol. PM180. SPIE Press, Bellingham, Washington, USA, 4th ed., 2008. http://spie.org/Publications/Book/781851, Accessed 2015-02-26.

[82] W. T. Welford, *Aberrations of optical systems*. CRC Press, Taylor & Francis Group, London, UK, 1986.

[83] Schott Inc., "TIE-29: Refractive Index and Dispersion," 2014. http://www.us.schott.com/advanced_optics/english/download/schott-tie-29-refractive-index-and-dispersion-april-2014-us.pdf, Accessed 2015-01-28.

[84] Schott Inc., "TIE-19: Temperature Coefficient of the Refractive Index," 2008. http://lightmachinery.com/media/1552/schott_tie-19_temperature_coefficient_of_refractive_index.pdf, Accessed 2015-01-28.

[85] W. J. Smith, *Modern lens design*, vol. 2. McGraw-Hill, New York, USA, 2005.

[86] Schott Inc., "TIE-31: Mechanical and thermal properties of optical glass," 2004. http://www.us.schott.com/advanced_optics/english/download/schott_tie-31_mechanical_and_thermal_properties_of_optical_glass_july_2004-_eng.pdf, Accessed 2015-01-28.

[87] P. R. Yoder, *Opto-Mechanical Systems Design*. Optical Science and Engineering, SPIE Press, Bellingham, Washington, USA, 3rd ed., 2006.

[88] Schott Inc., "TIE-27: Stress in optical glass," July 2004. http://fp.optics.arizona.edu/optomech/references/glass/Schott/tie-27_stress_in_optical_glass_us.pdf, Accessed 2015-01-27.

[89] A. K. Spilman and T. G. Brown, "Stress-induced focal splitting," *Optics express*, vol. 15, no. 13, pp. 8411–8421, 2007.

[90] S. M. Pompea, "Assessment of black and spectrally selective surfaces for stray light reduction in telescope systems," vol. 7739, p. 773921, SPIE, 2010.

[91] A. Duparré, "Characterization of surface and thin-film roughness using PSD functions," in *Optical Fabrication and Testing*, p. OTuA5, Optical Society of America, 2008.

[92] H. Gross, M. Hofmann, R. Jedamzik, P. Hartmann, and S. Sinzinger, "Measurement and simulation of striae in optical glass," vol. 7389, pp. 73891C–73891C–9, 2009. http://dx.doi.org/10.1117/12.827677.

[93] Schott Inc., "TIE-26: Homogeneity of Optical Glass," July 2004. http://fp.optics.arizona.edu/optomech/references/glass/Schott/tie-26_homogeneity_of_optical_glass_us.pdf, Accessed 2015-01-26.

[94] Schott Inc., "TIE-28: Bubbles and Inclusions in Optical Glass," 2004. http://fp.optics.arizona.edu/optomech/references/glass/Schott/tie-28_bubbles_and_inclusions_us.pdf, Accessed 2015-01-27.

[95] D. Golini, W. I. Kordonski, P. Dumas, and S. J. Hogan, "Magnetorheological finishing (MRF) in commercial precision optics manufacturing," vol. 3782, pp. 80–91, 1999. http://dx.doi.org/10.1117/12.369174.

[96] I. Iordanoff, A. Battentier, J. Néauport, and J. Charles, "A discrete element model to investigate sub-surface damage due to surface polishing," *Tribology International*, vol. 41, no. 11, pp. 957 – 964, 2008. http://dx.doi.org/10.1016/j.triboint.2008.02.018.

[97] J. E. Harvey, J. Lentz, and J. Joseph B. Houston, ""Just-Good-Enough" Optical Fabrication," in *Optical Fabrication and Testing*, p. OWA5, Optical Society of America, 2008.

[98] J. J. Guregian, R. T. Benoit, and W. K. Wong, "Overview of contamination effects on the performance of high-straylight-rejection telescopes via ground measurements," in *Proc. SPIE* (A. P. M. Glassford, ed.), vol. 1329, pp. 2–15, SPIE, 1990.

[99] K. A. Herren and D. A. Gregory, "Mie scattering of growing molecular contaminants," *Optical Engineering*, vol. 46, no. 3, p. 033602, 2007.

[100] A. C. Tribble, *Fundamentals of contamination control*, vol. 44. SPIE Press, Bellingham, Washington, USA, 2000.

[101] J. Wertz and W. Larson, *Space Mission Analysis and Design*. Space Technology Library, Springer, Berlin, Germany, 1999.

[102] R. E. Fischer, B. Tadic-Galeb, P. R. Yoder, and R. Galeb, *Optical system design*. McGraw Hill, New York, USA, 2000.

[103] D. Ristau, *Laser-induced damage in optical materials*. CRC Press, Taylor & Francis Group, London, UK, 2015.

[104] D. B. Doyle and R. H. Czichy, "Influences of simulated space radiation on optical glasses," in *Proc. SPIE*, vol. 2210, pp. 434–448, 1994. http://dx.doi.org/10.1117/12.188103.

[105] Schott Inc., "TIE-42: Radiation Resistant Optical Glasses," August 2007. http://www.us.schott.com/advanced_optics/english/download/schott_tie-42_radiation_resistant_optical_glasses_august_2007_eng.pdf, Accessed 2015-01-28.

[106] D. Bates, "Rayleigh scattering by air," *Planetary and Space Science*, vol. 32, no. 6, pp. 785 – 790, 1984. http://dx.doi.org/10.1016/0032-0633(84)90102-8.

[107] D. A. Long, *Raman spectroscopy*, vol. 206. McGraw-Hill, New York, USA, 1977.

[108] K. Kneipp, M. Moskovits, and H. Kneipp, *Surface-enhanced Raman scattering: physics and applications*, vol. 103. Springer, Berlin, Germany, 2006.

[109] E. Hecht, *Optics*. Addison-Wesley, Boston, MA, USA, 4th ed., 2002.

Chapter 6

Journal Paper Submission

The paper submitted to an accredited journal for publication is reproduced below. The journal in question is Current Applied Physics, an Elsevier publication. The subject matter of the paper arose from consideration of how to model image formation in the presence of a phase retarder implemented as a tilted plano-parallel glass plate in the pupil of an imaging system. This phase retarder was introduced in Chapter 3 as a type of segmented pupil interferometer that could be of assistance when measuring weak optical scatter near the specular beam. The technique could also be of use in performing retrieval of optical phase or aberration information when evaluating the image plane performance of optical systems. Considered in other contexts and more generally, this type of circular-pupil phase retarder is better known as the fractional Hilbert mask.

The acceptance correspondence and Article in Press are provided in Appendix B.

Zernike-basis expansion of the fractional and radial Hilbert phase masks

N Chetty[1] and D J Griffith[1]

[1] *School of Chemistry and Physics, University of KwaZulu-Natal (PMB)*
*Private Bag X01, Scottsville 3209, South Africa**

compiled: October 2, 2014

The linear Hilbert phase mask or transform has found applications in image processing and spectroscopy. An optical version of the fractional Hilbert mask is considered here, comprising an imaging system with a circular, unobscured pupil in which a variable phase delay is introduced into one half of the pupil, split bilaterally.
The radial Hilbert phase mask is also used in image processing and to produce optical vortices which have applications in optical tweezers and the detection of exoplanets.
We subjected the fractional and radial Hilbert phase masks to Zernike function expansion in order to compute the image plane electromagnetic field distribution using Nijboer-Zernike theory. The Zernike functions form an orthogonal basis on the unit circle. The complex-valued Zernike expansion coefficients for these two phase masks were derived for use in the context of the Extended Nijboer-Zernike (ENZ) theory of image formation. The ENZ approach is of interest in that it allows a greater range of defocus to be dealt with, provides a simple means of taking a finite source size into account and has been adapted to high Numerical Aperture (NA) imaging applications.
Our image plane results for the fractional Hilbert mask were verified against a numerical model implemented in the commercial optical design and analysis code, Zemax®. It was found that the Nijboer-Zernike result converged to the Zemax® result from below as the number of Zernike terms in the expansion was increased.

1. Introduction

Many problems in optics relate to the diffraction propagation of electro-magnetic fields. In the modeling of imaging systems, computation of the shape of the wavefront emerging from the system exit pupil is often fairly straightforward and can be accomplished to adequate accuracy using simple geometrical raytracing. The final step, that of diffraction propagation of the emergent wavefront into the focal region is then the key computation in arriving at the complex field amplitude and hence the radiant intensity at or near focus. This involves evaluation of a diffraction integral, some of the most general being the Rayleigh-Sommerfeld integrals [1].

This technique of using geometrical optics to map the field in the exit pupil followed by numerical evaluation of a diffraction integral is one of the methods used in optical design and analysis software codes. The diffraction integral that is numerically evaluated can be selected to suit the problem. Various approximations can be used to simplify the most general diffraction integrals, including the Debye, Fresnel and Fraunhofer approximations [1]. The Fresnel and Fraunhofer approximations are of special significance in that diffraction propagation can be expressed in terms of linear and shift-invariant operations in frequency space giving rise to the domain of Fourier optics [2]. This permits application of all the analytical and numerical Fourier techniques to a broad range of optical diffraction problems.

Wavefront aberrations can be expanded in terms of the Zernike functions, which form an orthogonal set of basis functions on the unit circle. This covers the case of the unobscured, circular pupil which is very common in imaging systems. Not only does Zernike-basis expansion aid in the diagnostic interpretation of test interferograms, for example, it also allows for application of the Classical Nijboer-Zernike (CNZ) theory in optical analysis [1]. In CNZ theory, each Zernike function mode in the exit pupil gives rise to a proportional field contribution in the image plane [1].

More recently [3], the Extended Nijboer-Zernike (ENZ) theory has been introduced permitting coverage of a broader range of diffraction imaging phenomena. A detailed description of the ENZ approach with extensive references has been made available in [4]. Here, we apply the ENZ theory to analyse the circular pupil, fractional Hilbert mask [5] and perform a numerical verification of the result. ENZ expansion coefficients for the radial Hilbert phase mask on a circular pupil are also provided together with some illustrations of the Point Spread Function (PSF) in the focal region.

2. The Zernike Functions

The normalized Zernike functions $\hat{Z}_n^m(\rho)$ are composed of the product of a normalization factor N_n^m, a radial Zernike polynomial $R_n^m(\rho)$ and azimuthal sine or cosine factors for $n \geq 0$ and $m = -n, -n+2, \ldots, n-2, n$ as [6]

* ChettyN3@ukzn.ac.za

$$\hat{Z}_n^m(\rho,\theta) = \begin{cases} N_n^m R_n^m(\rho)\cos m\theta & m \geq 0 \\ -N_n^m R_n^{|m|}(\rho)\sin m\theta & m < 0. \end{cases} \quad (1)$$

The restriction on m to vary from $-n$ to n in increments of 2 implies that if n is odd, m is also odd and if n is even, m will be even or zero. Conversely, if m is odd then n is also odd and if m is even or zero then n is also even or zero.

Normalization of the Zernike functions is not always included in the definition (Equation 1) by authors (excluded by Born and Wolf [1], included by Noll [7] and Thibos *et al.* [6] for example). If the normalization factors are excluded from the Zernike function definition, the normalization factors are then effectively incorporated into the expansion coefficients (denoted α_n^m in the following section).

While the product of the radial polynomial and cosine/sine factors always varies from -1 to $+1$ [8, 9], the normalization factor N_n^m for a specific Zernike function is given by [6, 7]

$$N_n^m = \sqrt{\frac{2(n+1)}{1+\delta_{m,0}}}, \quad (2)$$

where the Kronecker delta $\delta_{m,n}$ is

$$\delta_{m,n} = \begin{cases} 1 & m = n \\ 0 & m \neq n. \end{cases} \quad (3)$$

The radial Zernike polynomials can be written explicitly [1] as

$$R_n^m(\rho) = \sum_{s=0}^{p} \frac{(-1)^s (n-s)!}{s!(p-s)!(q-s)!}\rho^{n-2s} \quad (4)$$

where $p \equiv {}^{(n-|m|)}/_2$ and $q \equiv {}^{(n+|m|)}/_2$ are integers that are always zero or positive and the exclamation mark (!) denotes the factorial of the preceding expression. The radial polynomial coefficients are all integers which always sum to unity (i.e. $R_n^m(1) = 1$ for all m and n).

2.A. Zernike Function Orthogonality

The Zernike functions as defined in Equation 1 obey a set of inner product orthogonality relations [1]. The inner product of a function $\Phi(\rho,\theta)$ with the Zernike functions is defined as

$$\langle \Phi(\rho,\theta), \hat{Z}_n^m(\rho,\theta)\rangle \equiv \frac{1}{\pi}\int_0^{2\pi}\int_0^1 \Phi(\rho,\theta)\hat{Z}_n^m(\rho,\theta)\rho d\rho d\theta. \quad (5)$$

The inner products of the Zernike functions with themselves are

$$\langle \hat{Z}_n^m(\rho,\theta), \hat{Z}_{n'}^{m'}(\rho,\theta)\rangle = \delta_{m,m'}\delta_{n,n'}. \quad (6)$$

For the radial polynomials

$$\int_0^1 R_n^m(\rho)R_{n'}^m(\rho)\rho d\rho = \frac{\delta_{n,n'}}{2(n+1)}, \quad (7)$$

and

$$\int_0^{2\pi} \cos m\theta \cos m'\theta d\theta = \pi(1+\delta_{m,0})\delta_{m,m'} \quad (8)$$

in the azimuthal direction. A similar relationship exists for the sine functions. Orthonormality implies that the integral of the product of any two Zernike functions over the unit circle is zero unless the two functions are identical (have the same n and m). In the latter case, the integral is unity.

Since the Zernike functions form a complete, orthogonal basis on the unit circle, any (pupil) function $\Phi(\rho,\theta)$ defined on the unit circle can be expanded in terms of the Zernike functions. The normalized Zernike coefficients $\hat{\alpha}_n^m$ in the Zernike representation of a function are unique for for a given pupil function, regardless of how many terms are fitted. It also means that the coefficients (normalized in the following instance) for reconstructing $\Phi(\rho,\theta)$ can be found using the inner product as

$$\hat{\alpha}_n^m = \langle \Phi(\rho,\theta), \hat{Z}_n^m(\rho,\theta)\rangle. \quad (9)$$

If the Zernike functions are not normalized by definition, the orthogonality relationship in Equation 6 will be

$$\langle Z_n^m(\rho,\theta), Z_{n'}^{m'}(\rho,\theta)\rangle = \frac{1+\delta_{m,0}}{2(n+1)}\delta_{m,m'}\delta_{n,n'} \quad (10)$$

$$= \frac{1}{N_n^m N_{n'}^{m'}}\delta_{m,m'}\delta_{n,n'}. \quad (11)$$

2.B. Properties and Relationships of the Zernike Radial Polynomials

The Zernike functions and radial polynomials have been studied in detail [9, 10] in the general context of orthogonal polynomials [11]. Some special values and identities of the Zernike radial polynomials that will be required are [10]

$$R_n^m(1) = 1, \quad (12)$$

$$R_n^m(0) = (-\delta_{m,0})^{n/2}, \quad (13)$$

$$\int_0^1 R_n^m(\rho)d\rho = \frac{(-1)^p}{n+1}, \quad (14)$$

$$\int_0^1 R_n^m(\rho)\rho d\rho = \begin{cases} \frac{1}{2} & m,n = 0 \\ \frac{(-1)^p m}{n(n+2)} & \text{otherwise.} \end{cases} \quad (15)$$

Equations 14 and 15 could not be found explicitly in existing literature and are therefore proven here. Starting with a Zernike polynomial relationship provided by Prata and Rusch [12] and having defined

$$S_n^m(s,\rho) \equiv \int R_n^m(\rho)\rho^s d\rho, \tag{16}$$

they provide

$$\begin{aligned} S_n^m(0,\rho) = \frac{1}{n+1}[&(R_{n+1}^{m+1}(\rho) - R_{n+1}^{m+3}(\rho) + \ldots) \\ &- (R_{n-1}^{m+1}(\rho) - R_{n-1}^{m+3}(\rho) + \ldots)], \end{aligned} \tag{17}$$

where the summations stop when the azimuthal (upper) index exceeds the radial index. This is now rewritten using summation notation as

$$\begin{aligned} S_n^m(0,\rho) = &\frac{1}{n+1}\sum_{k=m+1,m+3\ldots}^{n+1}(-1)^{\frac{k-m-1}{2}}R_{n+1}^k(\rho) \\ &-\frac{1}{n+1}\sum_{k=m+1,m+3\ldots}^{n-1}(-1)^{\frac{k-m-1}{2}}R_{n-1}^k(\rho). \end{aligned} \tag{18}$$

From Equation 16, the definite integral

$$\int_0^1 R_n^m(\rho)d\rho = S_n^m(0,\rho) - S_n^m(0,0) = S_n^m(0,1), \tag{19}$$

since from Equation 18 with Equation 13 it is clear that $S_n^m(0,0) = 0$. Then substituting $\rho = 1$ in Equation 18 and using the identity in Equation 12 yields

$$\begin{aligned} S_n^m(0,1) = &\frac{1}{n+1}\sum_{k=m+1,m+3\ldots}^{n+1}(-1)^{\frac{k-m-1}{2}} \\ &-\frac{1}{n+1}\sum_{k=m+1,m+3\ldots}^{n-1}(-1)^{\frac{k-m-1}{2}}. \end{aligned} \tag{20}$$

The summations cancel term for term, except for the final term with $k = n+1$ in the first summation, so that

$$S_n^m(0,1) = \frac{(-1)^{\frac{n-m}{2}}}{n+1} = \frac{(-1)^p}{n+1}, \tag{21}$$

thus proving Equation 14.

Prata and Rusch [12] show that

$$\begin{aligned} S_n^m(1,\rho) = \frac{1}{2(n+1)}[&(n+m+2)S_{n+1}^{m+1}(0,\rho) \\ &+ (n-m)S_{n-1}^{m+1}(0,\rho)]. \end{aligned} \tag{22}$$

From Equation 16, the definite integral

$$\int_0^1 R_n^m(\rho)\rho d\rho = S_n^m(1,\rho) - S_n^m(1,0) = S_n^m(1,1). \tag{23}$$

By substituting $\rho = 0$ in Equation 22 it is clear that $S_n^m(1,0) = 0$. Then substituting $\rho = 1$ in Equation 22 and using Equation 21,

$$\begin{aligned} S_n^m(1,1) = &\frac{1}{2(n+1)}[(n+m+2)\frac{(-1)^{\frac{n-m}{2}}}{n+2} \\ &+ (n-m)\frac{(-1)^{\frac{n-m-2}{2}}}{n}] \end{aligned} \tag{24}$$

$$\begin{aligned} = &\frac{1}{2(n+1)}[(n+m+2)\frac{(-1)^p}{n+2} \\ &- (n-m)\frac{(-1)^p}{n}] \end{aligned} \tag{25}$$

$$= \frac{(-1)^p}{2(n+1)}\left[\frac{n+m+2}{n+2} - \frac{n-m}{n}\right] \tag{26}$$

$$= \frac{(-1)^p m}{n(n+2)}. \tag{27}$$

The final step is some straightforward algebra, thus proving Equation 15, except for the case $m = n = 0$, follows easily from $R_0^0(\rho) = 1$.

3. Fractional Hilbert Mask Expansion

The goal here is to find the ENZ coefficients of the one-dimensional, fractional Hilbert mask on a circular, unobscured pupil. Optically, this is equivalent to splitting the pupil bilaterally and introducing a variable phase shift into one half.

The Zernike coefficients are computed using the inner product as given in Equation 9 and the inner product definition in Equation 5. The ENZ approach differs with reference to the classical Nijboer-Zernike (CNZ) approach in that the complex-valued EM pupil field amplitude is expanded instead of the real-valued wavefront error. In the ENZ case, the formalism is essentially the same as for the classical approach, except that the Zernike coefficients are now generally also complex-valued [4] and are usually denoted β_n^m. We will use A_n^m and B_n^m for the real and complex-valued Zernike coefficients for the circular pupil fractional Hilbert mask.

For a general, complex-valued pupil function $P(\rho,\theta)$ the normalized ENZ expansion is (using the inner product notation defined in Equation 5)

$$\hat{\beta}_n^m = \langle P(\rho,\theta), \hat{Z}_n^m(\rho,\theta)\rangle. \tag{28}$$

Note that in what follows, the ENZ expansions will only consider cosine terms ($m \geq 0$), and therefore any modulus signs around m will be dropped.

To perform the ENZ expansion of the circular fractional Hilbert mask, the circular pupil is first segmented bilaterally by expressing the (real-valued) pupil transmission $A(\theta)$ in the azimuthal coordinate, θ, as an infinite Fourier cosine series representing a square wave [13–15] as

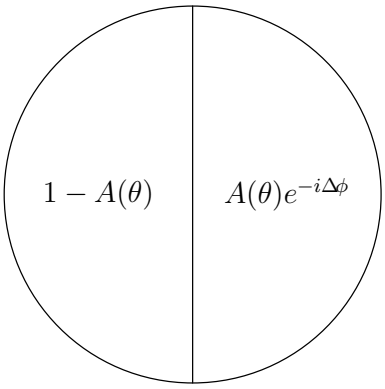

Figure 1. Fractional Hilbert Mask Pupil Function

$$A(\theta) = \frac{1}{2} + \frac{2}{\pi}\sum_{k=1}^{\infty}\frac{\sin\frac{k\pi}{2}}{k}\cos k\theta \tag{29}$$

$$= \frac{1}{2} + \frac{2}{\pi}\sum_{k=1,3,5\ldots}^{\infty}\frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta. \tag{30}$$

A Fourier series of this kind, when truncated, exhibits the Gibbs phenomenon (ripple and overshoot near discontinuities [14]) and requires many terms in the summation for reasonable convergence. This aspect will discussed in more detail in §3.A.

The real transmission function $A(\theta)$ is ideally unity in the right half-plane ($x > 0$) and zero in the left half-plane ($x < 0$). So for the Hilbert mask, the right half of the pupil is expressed in transmission as $A(\theta)$, and the left half of the pupil is then $1 - A(\theta)$.

The full, complex-valued pupil with a phase retardation of $\Delta\phi$ in the right half is then written as

$$\begin{aligned} P(\rho,\theta) &= A(\theta)e^{-i\Delta\phi} + 1 - A(\theta) \\ &= (e^{-i\Delta\phi} - 1)A(\theta) + 1. \end{aligned} \tag{31}$$

This is illustrated in Figure 1.

The real-valued transmission function $A(\theta)$ is subject to normalized (coefficients for normalized Zernike functions are denoted $\hat{A}_n^m$, $\hat{B}_n^m$, $\hat{\alpha}_n^m$ etc.) Zernike expansion first as

$$\begin{aligned} \hat{A}_n^m &= \langle A(\theta), \hat{Z}_n^m(\rho,\theta)\rangle \\ &= \frac{1}{2}\langle \hat{Z}_0^0, \hat{Z}_n^m(\rho,\theta)\rangle \\ &+ \frac{2}{\pi}\langle \sum_{k=1,3,5\ldots}^{\infty}\frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta, \hat{Z}_n^m(\rho,\theta)\rangle \\ &= \frac{1}{2}\langle \hat{Z}_0^0, \hat{Z}_n^m(\rho,\theta)\rangle \\ &+ \frac{2N_n^m}{\pi}\langle \sum_{k=1,3,5\ldots}^{\infty}\frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta, R_n^m(\rho)\cos m\theta\rangle. \end{aligned} \tag{32}$$

The final inner product in Equation 32 is evaluated using the definition in Equation 6 as

$$\begin{aligned} &\langle \sum_{k=1,3,5\ldots}^{\infty}\frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta, R_n^m(\rho)\cos m\theta\rangle \\ &= \frac{1}{\pi}\int_0^{2\pi}\int_0^1\left[\sum_{k=1,3,5\ldots}^{\infty}\frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta\right]R_n^m(\rho)\cos m\theta\rho d\rho d\theta \\ &= \frac{1}{\pi}\int_0^{2\pi}\left[\sum_{k=1,3,5\ldots}^{\infty}\frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta\right]\cos m\theta d\theta\int_0^1 R_n^m(\rho)\rho d\rho \\ &= \frac{1}{\pi}\sum_{k=1,3,5\ldots}^{\infty}\frac{(-1)^{\frac{k-1}{2}}}{k}\int_0^{2\pi}\cos k\theta\cos m\theta d\theta\int_0^1 R_n^m(\rho)\rho d\rho. \end{aligned} \tag{33}$$

Applying the cos-function orthogonality relationship given in Equation 8 as well as Equation 15 we notice that all terms in the summation vanish except for when $k = m$. Thus

$$\begin{aligned} &\langle \sum_{k=1,3,5\ldots}^{\infty}\frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta, R_n^m(\rho)\cos m\theta\rangle \\ &= \frac{1}{\pi}\sum_{k=1,3,5\ldots}^{\infty}\frac{(-1)^{\frac{k-1}{2}}}{k}\pi\delta_{k,m}\int_0^1 R_n^m(\rho)\rho d\rho \\ &= \begin{cases} \frac{(-1)^{\frac{m-1}{2}}}{m}\frac{(-1)^{\frac{n-m}{2}}m}{n(n+2)} & m,n \text{ positive, odd} \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{(-1)^{\frac{n-1}{2}}}{n(n+2)} & m,n \text{ positive, odd} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{34}$$

Substituting Equation 34 in Equation 32 gives

$$\hat{A}_n^m = \begin{cases} \frac{1}{2} & m = n = 0 \\ \frac{2N_n^m(-1)^{\frac{n-1}{2}}}{\pi n(n+2)} & m,n \text{ odd} \\ 0 & m,n \text{ even.} \end{cases} \tag{35}$$

It is noted that these coefficients are independent of m (the coefficients are the same across all values of m for each n).

The fractional Hilbert mask pupil function $P(\rho,\theta)$ is then expanded using the ENZ procedure as

$$\begin{aligned} \hat{B}_n^m &= \langle P(\rho,\theta), \hat{Z}_n^m(\rho,\theta)\rangle \\ &= \langle (e^{-i\Delta\phi} - 1)A(\theta) + 1, \hat{Z}_n^m(\rho,\theta)\rangle \\ &= (e^{-i\Delta\phi} - 1)\hat{A}_n^m + \langle \hat{Z}_0^0, \hat{Z}_n^m(\rho,\theta)\rangle. \end{aligned} \tag{36}$$

The $\hat{B}_n^m$ coefficients can thus be written down from the $\hat{A}_n^m$ coefficients (Equation 35) as

$$\hat{B}_n^m = \begin{cases} \frac{(e^{-i\Delta\phi}+1)}{2} & m=n=0 \\ \frac{2N_n^m(e^{-i\Delta\phi}-1)(-1)^{\frac{n-1}{2}}}{\pi n(n+2)} & m,n \text{ odd} \\ 0 & m,n \text{ even.} \end{cases} \quad (37)$$

This solution for $\hat{B}_n^m$ has the required property that if $\Delta\phi \to 0$ then $\hat{B}_n^m \to 0$ for all positive m and n while $\hat{B}_0^0 \to 1$. These coefficients are also independent of m.

3.A. Gibbs Phenomenon

Formulation of the fractional Hilbert mask pupil function involved the use of a square wave, synthesized as a Fourier cosine sequence $A(\theta)$ given in Equation 29. This series has relatively poor convergence and exhibits ripple (or ringing) and overshoot artifacts near the discontinuities, known as "Gibbs Phenomenon" [14] when the series is truncated in numerical work. It follows that these artifacts will be propagated by the numerical ENZ process to the focal region results when the series is necessarily truncated.

The Gibbs phenomenon can be suppressed to a significant extent (but not eliminated) using a filtering factor called the Lanczos σ-factor [14]. If the series is to be truncated at $n = N$, the Fourier coefficients are multiplied by $\mathrm{sinc}\frac{n}{N}$ where $\mathrm{sinc}(x) \equiv {}^{\sin \pi x}/_{\pi x}$ and the summation is performed up to $n = N-1$. So truncating the series at $n = N$ and including the Lanczos σ-factor, the $\hat{B}_n^m$ coefficients for normalized Zernike functions become

$$\hat{B}_n^m = \begin{cases} \frac{e^{-i\Delta\phi}+1}{2} & m=n=0 \\ \frac{2N_n^m(e^{-i\Delta\phi}-1)(-1)^{\frac{n-1}{2}}\mathrm{sinc}\frac{n}{N}}{\pi n(n+2)} & m,n \text{ odd, } n<N \\ 0 & \text{otherwise.} \end{cases} \quad (38)$$

The corresponding result for the $\hat{A}_n^m$, being the normalized Zernike coefficients for $A(\theta)$ follows from Equation 35 as

$$\hat{A}_n^m = \begin{cases} \frac{1}{2} & m=n=0 \\ \frac{2N_n^m(-1)^{\frac{n-1}{2}}\mathrm{sinc}\frac{n}{N}}{\pi n(n+2)} & m,n \text{ odd, } n<N \\ 0 & \text{otherwise.} \end{cases} \quad (39)$$

As an illustration, $A(\theta)$ is plotted for a summation of 50 terms with and without the Lanczos σ-factor in Figure 2. Clearly the inclusion of the σ-factor helps to suppress the Gibbs phenomenon. The first discontinuity is at $\theta = \pi/2$.

With or without the Lanczos σ-factor for suppression of the Gibbs phenomenon, it is necessary to compute the Zernike radial polynomials to high order for numerical verification purposes. The coefficients (see Equation 4) are ratios of factorials which become large, computationally expensive and potentially inaccurate from $n = 40$ onward [16]. Alternative techniques for computation of the polynomial coefficients are then required. Janssen and Dirksen [16] and Vlcek and Sovka [17] have provided solutions which exploit the relationship of the Zernike radial polynomials to Chebyshev polynomials and the more general Jacobi polynomials.

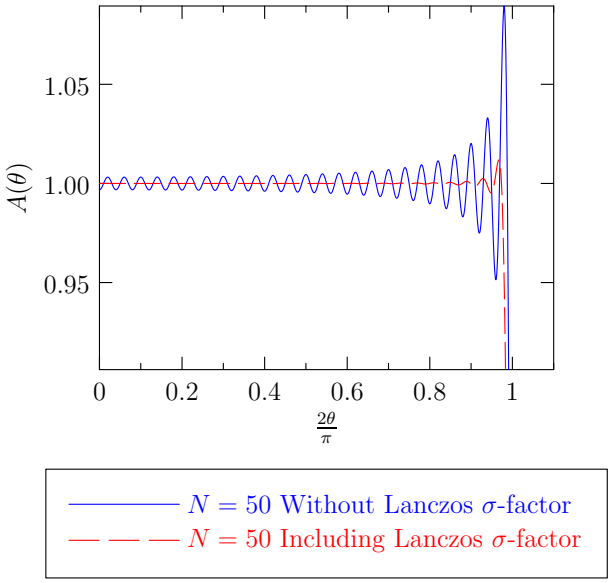

Figure 2. $A(\theta)$ Truncated at 50 Terms

4. ENZ Propagation from Pupil to Focal Region

The essence of the ENZ approach [3] is that each Zernike term in the complex-valued pupil expansion corresponds to a field contribution in the focal region. The complex-valued, scalar field amplitude, U, in the focal region is written as

$$U(\rho', \theta'; d_f) = \sum_{m,n} \beta_n^m U_n^m(\rho', \theta'; d_f), \quad (40)$$

where $\beta_n^m U_n^m$ is the contribution from $\beta_n^m Z_n^m$ in the exit pupil, the lateral coordinates in the focal region are (ρ', θ'), with a relative defocus from the best focal plane in the axial direction of d_f. Janssen [3] evaluated the focal region contributions U_n^m for substantial defocus d_f (relative to CNZ), by deriving functions $V_n^m(\rho', d_f)$ such that

$$U_n^m(\rho', \theta'; d_f) = 2i^m V_n^m(\rho', d_f) \cos m\theta', \quad (41)$$

where the radial and defocus function V_n^m is an integral over Zernike polynomial, radially-weighted Bessel functions, J_m of the first kind and order m,

$$V_n^m(\rho', d_f) = \int_0^1 \exp\left\{id_f\rho^2\right\} R_n^m(\rho) J_m(2\pi\rho\rho')\rho d\rho. \quad (42)$$

Note that the focal region coordinates and defocus parameter d_f are not in absolute length units here, but

expressed relative to the Numerical Aperture (NA) of imaging and the wavelength of light under consideration. Exit pupil and focal region coordinates are illustrated in §4.A.

Power-Bessel series expansion on V_n^m produces [3]

$$V_n^m(\rho', d_f) = e^{id_f} \sum_{l=1}^{\infty} (-2id_f)^{l-1} \sum_{j=0}^{p} v_{lj} \frac{J_{m+l+2j}(2\pi\rho')}{l\,(2\pi\rho')^l}, \tag{43}$$

in which $p = {}^{(n-m)}/_2$ (as before for positive m). The v_{lj} factor is computed as

$$v_{lj} = (-1)^p (m+l+2j) \binom{m+j+l-1}{l-1} \binom{j+l-1}{l-1} \times \binom{l-1}{p-j} / \binom{q+l+j}{l}, \tag{44}$$

with $q = {}^{(n+m)}/_2$ (as before for positive m).

For defocus parameter $d_f = 0$, the result of the Classical Nijboer-Zernike (CNZ) theory can be used in which the summation over l in Equation 43 is reduced to a single Bessel term as [4]

$$\begin{aligned} V_n^m(\rho', 0) &= \int_0^1 R_n^m(\rho) J_m(2\pi\rho\rho')\rho d\rho \\ &= (-1)^{\frac{n-m}{2}} \frac{J_{n+1}(2\pi\rho')}{2\pi\rho'}. \end{aligned} \tag{45}$$

The CNZ result in Equation 45 is useful for consistency checking and for the initial evaluation of convergence. Also, the evaluation of the v_{lj} binomial factors as given in Equation 44 can become numerically problematic at high order.

4.A. Pupil and Focal Region Coordinates

The radial coordinate ρ in the pupil is normalized relative to the physical pupil radius. The Cartesian coordinates (x', y', d_f) and polar coordinates (ρ', θ') in the focal region are also normalized relative to absolute coordinates (x, y, z) as follows:

$$\begin{aligned} x' &= x\frac{NA}{\lambda} = \rho' \cos\theta' \\ y' &= y\frac{NA}{\lambda} = \rho' \sin\theta' \\ d_f &= 2\frac{\pi}{\lambda} z(1 - \sqrt{1 - NA^2}), \end{aligned} \tag{46}$$

where NA is the Numerical Aperture (NA) of imaging and λ is the light wavelength.

The relative coordinate scheme is illustrated in Figure 3. Here the system exit pupil P with absolute outer edge radius R and relative radial coordinate ρ, has a spherical wavefront emerging from the exit pupil and converging on the image plane origin at O.

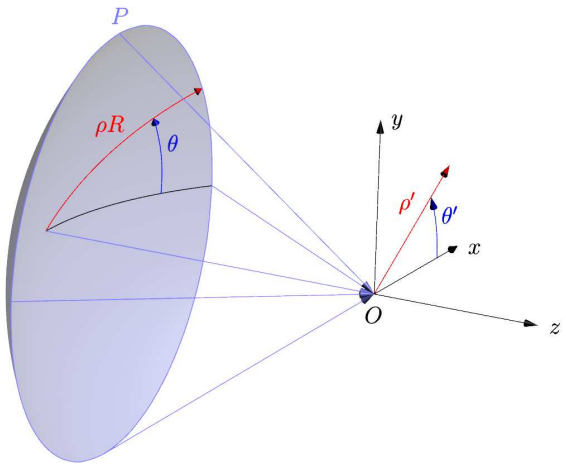

Figure 3. Exit Pupil (P) and Focal Region (O) Coordinates

4.B. Zernike Function Normalization in ENZ Practice

Braat *et al.* [4] exclude the normalisation factor N_n^m in their definition of the Zernike functions . In that case, the expressions for the unnormalized coefficients A_n^m and B_n^m for the fractional Hilbert mask (with reference to Equations 39 and 38, including the Lanczos σ-factor) will include the square of the N_n^m factors and

$$A_n^m = \begin{cases} \frac{1}{2} & m = n = 0 \\ \frac{2[N_n^m]^2 (-1)^{\frac{n-1}{2}} \operatorname{sinc}\frac{n}{N}}{\pi n(n+2)} & m, n \text{ odd},\ n < N \\ 0 & \text{otherwise}, \end{cases} \tag{47}$$

with

$$B_n^m = \begin{cases} \frac{e^{-i\Delta\phi}+1}{2} & m = n = 0 \\ \frac{2[N_n^m]^2 (e^{-i\Delta\phi}-1)(-1)^{\frac{n-1}{2}} \operatorname{sinc}\frac{n}{N}}{\pi n(n+2)} & m, n \text{ odd},\ n < N \\ 0 & \text{otherwise}. \end{cases} \tag{48}$$

4.C. ENZ Scope and Limitations

The scope and limitations of the application of ENZ tools *circa* 2010 is covered in detail by Van Haver [18]. The accuracy of this ENZ result relative to more accurate (numerical) Rayleigh integral computations is expected to degrade in both the low and high NA regimes. For optical wavelengths on the order of 550 nm in the visible spectrum, the applicable range is taken to be $0.05 < NA < 0.6$. Other limitations relating to vector and associated polarization effects are assumed negligible in the current context.

5. Numerical Verification

5.A. CNZ Convergence

The convergence of the CNZ result for the circular pupil, fractional Hilbert mask (Equations 45, 41, 40 with

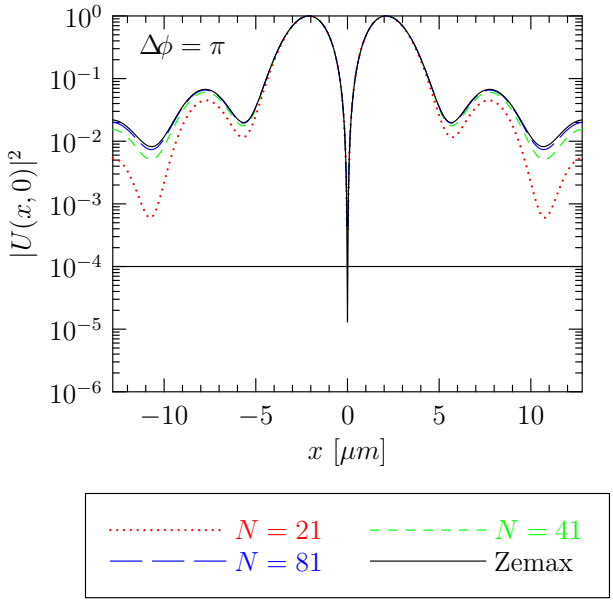

Figure 4. Circular Fractional Hilbert Pupil Mask,$\log_{10}$ Normalized Image PSF x-Cross Section (CNZ), $\Delta\phi = \pi$, $NA = 0.1$, $\lambda = 0.5\mu m$, $d_f = 0$

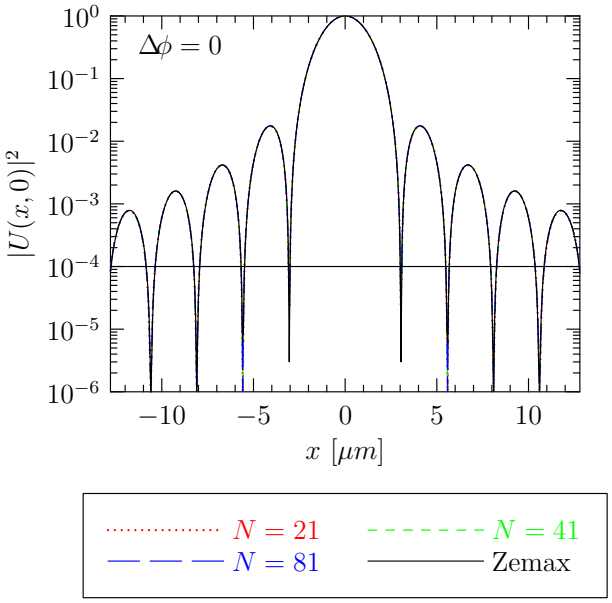

Figure 5. Circular Fractional Hilbert Pupil Mask,$\log_{10}$ Normalized Image PSF x-Cross Section (CNZ), $\Delta\phi = 0$, $NA = 0.1$, $\lambda = 0.5\mu m$, $d_f = 0$

complex-valued coefficients B_n^m computed with Equations 48 or 38) was tested with increasing N (the maximum value of n). A cross-section of the PSF intensity (irradiance) along the x-axis was computed for a sequence of increasing N with $\Delta\phi = \pi$, $NA = 0.1$ and $\lambda = 0.5\mu m$. The CNZ result without the normalization factor is plotted in Figure 4. This result is shown relative to the Non-Sequential Component (NSC) model result from the optical analysis package Zemax®. The Zemax® result will be discussed in more detail in §6.

A horizontal line at a relative irradiance which is 4 orders of magnitude below peak has been plotted in Figures 4. This is assumed to be measurable with high dynamic range Focal Plane Array (FPA) detectors.

The x-cross section convergence of the CNZ result relative to NSC Zemax® results for other values of $\Delta\phi$ are plotted in Figures 5, 6 and 7.

For $\Delta\phi = 0$ (Figure 5), there is only a single CNZ term, providing the normal incoherent diffraction-limited PSF for an unobscured, circular pupil (the Airy disk).

For non-zero $\Delta\phi$ convergence appears to be from below. The image-plane PSF for a number of different values of $\Delta\phi$ was computed using the CNZ result (Equation 45, $d_f = 0$) and the $\log_{10}$ of the normalized irradiance (square modulus of the complex-valued scalar field) is plotted in Figure 8 for $NA = 0.1$, $\lambda = 0.5\mu m$ and $N = 41$.

The result for $\Delta\phi = 0$ in Figure 8 yields the expected diffraction-limited PSF for $NA = 0.1$, $\lambda = 0.5\mu m$, corresponding to $m = n = 0$ in Equation 45. The full, central dark interference null is only formed close to $\Delta\phi = \pi$.

Braat *et al.* [4] provide Matlab® code for evaluation of the V_n^m functions, comprising implementation of Equation 43. This code was used for the computation of the field in the focal plane for the circular pupil, fractional Hilbert mask. The result was verified to be the same for the $d_f = 0$ case as for the CNZ result.

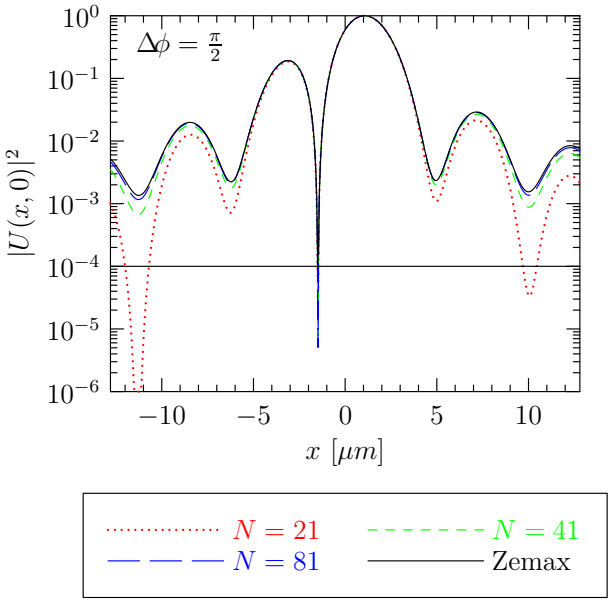

Figure 6. Circular Fractional Hilbert Pupil Mask, $\log_{10}$ Normalized Image PSF x-Cross Section (CNZ), $\Delta\phi = \pi/2$, $NA = 0.1$, $\lambda = 0.5\mu m$, $d_f = 0$

In the process it was noted that the radial coordinate input to the Matlab® function for V_n^m must be scaled by 2π in addition to normalizing as per Equation 46.

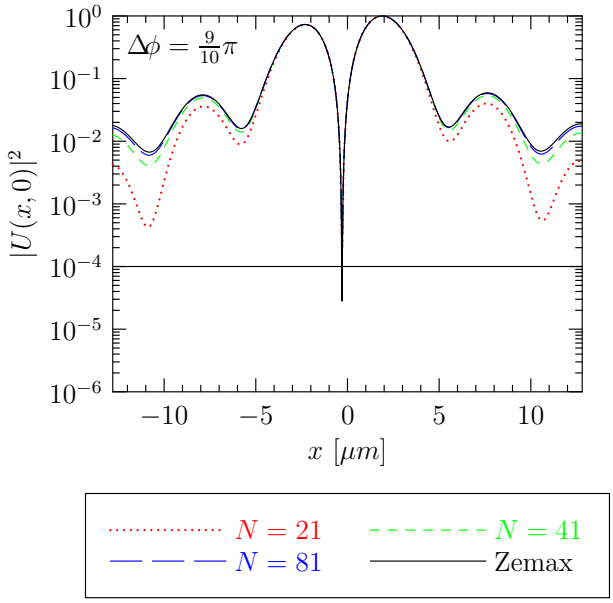

Figure 7. Circular Fractional Hilbert Pupil Mask, $\log_{10}$ Normalized Image PSF x-Cross Section (CNZ), $\Delta\phi = {}^{9\pi}/_{10}$, $NA = 0.1$, $\lambda = 0.5\mu m$, $d_f = 0$

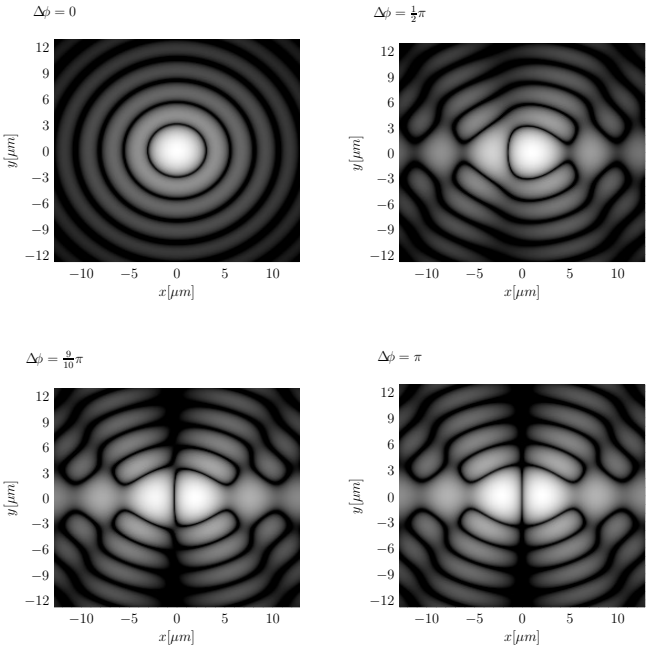

Figure 8. Circular Fractional Hilbert Pupil Mask, log Normalized Image PSF Irradiance (CNZ), $NA = 0.1$, $\lambda = 0.5\,\mu m$, $d_f = 0$, $N = 41$

6. Comparison to Zemax® Results

Zemax® is a multipurpose, optical design and analysis software code. Two methods of modeling the circular, fractional Hilbert mask problem were attempted. In the first method, the phase piston ($\Delta\phi$) was introduced into the optical layout using a Zemax® surface type called the "Zernike Standard Phase" surface. This surface directly introduces a phase advance/retardation into the wavefront through specification of the Zernike-basis coefficients. For various reasons (more detail is given in §6.A), this method did not work well and was useful only to help verify the Zernike coefficients $\hat{A}_n^m$ for normalized Zernike functions (Equation 39).

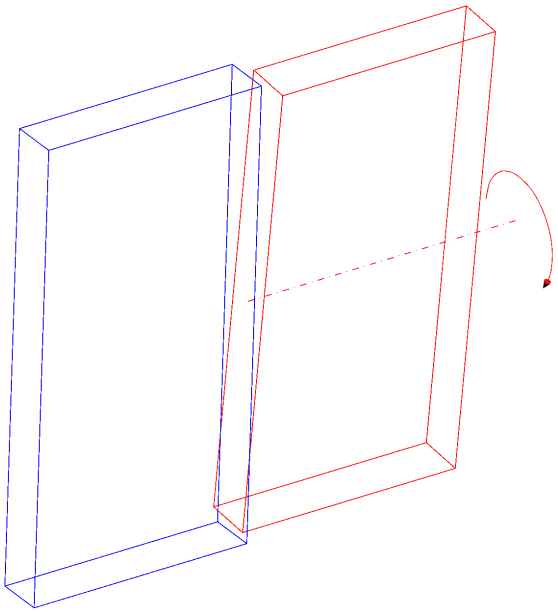

Figure 9. Tilted plane parallel glass plate phase retarder as fractional Hilbert mask implementation. The phase retarder would be placed in the system pupil.

The second Zemax® modeling method entailed introduction of the phase piston $\Delta\phi$ using a Non-Sequential Component (NSC) model of a tilted glass plate phase retarder as illustrated in Figure 9. It was necessary to use two pairs of plates in sequence, with one sequential pair tilted in opposite directions in order to undo the small lateral beam displacement caused by the first tilted plate.

Once the Zemax® NSC model of the plate retarder had been set up and verified, the cross-section of the PSF was computed, using direct summation of Huygens wavelets. This is thought to be the more accurate (and definitely more time-consuming) method with respect to the more routine FFT approach. Besides the Huygens wavelet summation and FFT methods, Zemax® also offers a Physical Optical Propagation (POP) method of analysis. The Zemax® result from the Huygens wavelet-summation method is shown as a reference in Figures 4, 5, 6 and 7. The CNZ result does converge to the Zemax result, but quite a large number of terms ($N > 80$) are required for close agreement.

6.A. Zernike Functions in Zemax

Zemax® follows the Noll [7] indexing scheme for the Zernike ("standard" as opposed to "fringe") functions which maps $Z_n^m \rightarrow Z_j$ according to sequence A176988 in the On-Line Encyclopedia of Integer Sequences. This mapping appears to lack the virtue of a simple formula, but follows the rule that odd j are assigned to values of $m < 0$ (and even j to all $m > 0$) and smaller values of j are assigned to smaller values of $|m|$.

Zemax only allows up to $j = 231$ Noll terms in the Zernike series. This corresponds to all terms up to $n = 21$. This is not really sufficient to allow for an accurate computation of the fractional Hilbert pupil mask situation using the Zernike Standard Phase surface in Zemax. Zemax® also follows Noll [7] in that the Zernike functions include the normalization factor.

6.B. Computation of Circular Pupil, Fractional Hilbert Mask PSF using Zemax®

The circular pupil, fractional Hilbert mask was modeled in Zemax® using a Non-Sequential Component (NSC) group and implemented as a tilted plate phase retarder illustrated in Figure 9. Zemax® currently offers three different methods of computing the PSF, namely the direct summation of Huygens wavelets, an FFT technique and a Physical Optics Propagation (POP) algorithm. All three methods produced essentially the same results illustrated in Figures 4 to 8.

The Huygens PSF results from Zemax® are visually indistinguishable from the results shown in Figure 8. Note that the x-axis cross sections of the Zemax® Huygens PSF image results are plotted as the reference curves in Figures 4, 5, 6 and 7.

7. The Radial Hilbert Phase Mask

The radial Hilbert phase mask is equivalent to a phase vortex used in [19] and in vortex coronagraph concepts for exoplanet detection [20]. In this case the pupil phase is written as

$$P(\theta) = e^{il\theta}, \tag{49}$$

where the integer vortex topological charge is $l \geq 1$. We performed unnormalised Zernike-basis expansion using the ENZ approach and for $m \geq 0$ as

$$\begin{aligned}
\beta_n^m &= \left\langle e^{il\theta}, Z_n^m(\rho,\theta)\right\rangle \\
&= \frac{1}{\pi}\int_0^{2\pi}\int_0^1 e^{il\theta} Z_n^m(\rho,\theta)\rho d\rho d\theta \\
&= \frac{1}{\pi}\int_0^{2\pi}\int_0^1 e^{il\theta} R_n^m(\rho)\cos m\theta \rho d\rho d\theta \\
&= \frac{1}{\pi}\int_0^{2\pi} e^{il\theta}\cos m\theta d\theta \int_0^1 R_n^m(\rho)\rho d\rho \\
&= \frac{1}{\pi}\int_0^{2\pi} (\cos l\theta + i\sin l\theta)\cos m\theta d\theta \int_0^1 R_n^m(\rho)\rho d\rho \\
&= \frac{\delta_{l,m}(-1)^p m}{n(n+2)}.
\end{aligned} \tag{50}$$

The final step arises from the orthogonality relations of the sin and cos functions and application of Equation 15. For $m < 0$, a similar argument yields

$$\beta_n^m = \frac{-i\delta_{l,|m|}(-1)^p|m|}{n(n+2)} = -i\beta_n^{-m}. \tag{51}$$

The vortex charge l selects the corresponding azimuthal orders $|m| = l$ only, but otherwise there are non-zero coefficients for all n. The ENZ focal region reconstruction proceeds according to Equation 40. Figure 10 shows the intensity and phase in focus ($d_f = 0$) and out of focus ($d_f = 3$ and $d_f = 6$).

8. Conclusion

The Zernike functions were reviewed as an orthogonal set of basis functions on the unit circle. The Zernike function basis has found numerous applications in optics and other disciplines such as image processing. A selection of Zernike function and Zernike radial polynomial relations were presented in the context of the Classical Nijboer-Zernike (CNZ) theory of optical aberrations. The more recent, and growing, Extended Nijboer-Zernike (ENZ) approach [3, 18] has expanded the range of optical diffraction problems that can be modeled and analyzed using an efficient semi-analytical procedure rooted in the use of Zernike-basis expansion.

The linear, fractional Hilbert mask has found applications in optics [5] and image processing [21]. The main result presented here was derivation of the Extended Nijboer-Zernike expansion coefficients for the circular pupil, fractional Hilbert mask. The convergence of the resulting infinite series, when truncated, was improved by introduction of the Lanczos σ-factor [14]. The ENZ coefficients for the circular pupil, fractional Hilbert mask were verified using numerical techniques, especially that of a comparison to results from the Zemax® optical design and analysis code.

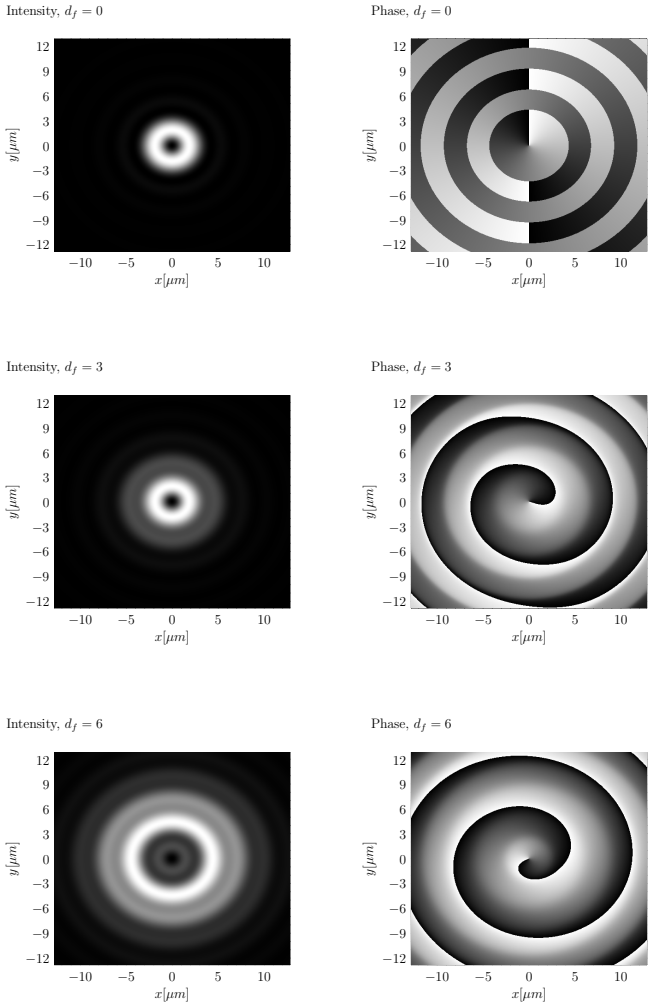

Figure 10. Optical vortex focal region normalized linear intensity and phase for charge $l = 1$. Normalised intensity varies from 0 (black) to 1 (white) and phase varies from 0 (black) to 2π(white), $NA = 0.1$, $\lambda = 0.5\,\mu m$.

The circular pupil ENZ expansion coefficients for the radial Hilbert phase mask, which is identified with the optical vortex, have also been provided together with illustrations of the focal region PSF intensity and phase.

Our future work in this area will relate to the use of Hilbert masks to produce axial irradiance nulls for high dynamic range measurements of scattering phase functions and also for aberration retrieval using irradiance measurements in the focal region.

References

[1] M. Born and E. Wolf, *Principles of Optics* (Cambridge University Press, 1999), 7th ed.

[2] J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, 1968).

[3] A. J. E. M. Janssen, "Extended Nijboer-Zernike approach for the computation of optical point-spread functions", J. Opt. Soc. Am. A **19**, 5, 849-857 (2002).

[4] J. J. M. Braat, P. Dirksen, S. van Haver and A. J. E. M. Janssen, "Detailed description of the ENZ approach", (2013), http://www.nijboerzernike.nl/.

[5] A. W. Lohmann, D. Mendlovic, and Z. Zalevsky, "Fractional Hilbert transform," Opt. Lett. **21**, 281-283 (1996).

[6] L. Thibos, R. A. Applegate, J. T. Schwiegerling, and R. Webb, "Standards for reporting the optical aberrations of eyes," in "Vision Science and its Applications," (Optical Society of America, 2000), p. SuC1.

[7] R. J. Noll, "Zernike polynomials and atmospheric turbulence", J. Opt. Soc. Am. **66**, 207-211 (1976).

[8] A. J. Janssen, "New analytic results for the Zernike circle polynomials from a basic result in the Nijboer-Zernike diffraction theory," Journal of the European Optical Society - Rapid publications **6** (2011).

[9] G. Szegö, *Orthogonal Polynomials*, Vol. 23 in American Mathematical Society colloquium publications (American Mathematical Society, 1939).

[10] W. J. Tango, "The circle polynomials of Zernike and their application in optics," Applied Physics **13**, 327-332 (1977).

[11] M. Abramowitz and I. A. Stegun, eds., "Handbook of Mathematical Functions," Vol. **55** of Applied Mathematics Series (NIST, 1964), 10th ed.

[12] J. Aluizio Prata and W. V. T. Rusch, "Algorithm for computation of Zernike polynomials expansion coefficients," Appl. Opt. **28**, 749-754 (1989).

[13] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*, (Dover Publications, 2001), 2nd ed., http://www-personal.umich.edu/~jpboyd/aaabook_9500may00.pdf.

[14] A. Jerri, *The Gibbs Phenomenon in Fourier Analysis, Splines and Wavelet Approximations*, Vol. 446 of Mathematics and Its Applications (Springer, 1998).

[15] K. Howell, *Principles of Fourier Analysis*, Studies in Advanced Mathematics (Taylor & Francis, 2010).

[16] A. Janssen and P. Dirksen, "Computing Zernike polynomials of arbitrary degree using the discrete Fourier transform," Journal of the European Optical Society - Rapid publications **2** (2007).

[17] M. Vlcek and P. Sovka, "Zernike polynomials and their spectral representation," in *Proceedings of the 2013 International Conference on Electronics, Signal Processing and Communication Systems*, (2013).

[18] S. Van Haver, "The Extended Nijboer-Zernike Diffraction Theory and its Applications," Ph.D. thesis (TU Delft, 2010), http://repository.tudelft.nl/view/ir/uuid:8d96ba75-24da-4e31-a750-1bc348155061/.

[19] D. G. Grier, "A revolution in optical manipulation," Nature 424, 810816 (2003).

[20] G. Foo, D. M. Palacios, and J. Grover A. Swartzlander, "Optical vortex coronagraph," Opt. Lett. 30, 33083310 (2005).

[21] J. A. Davis, D. E. McNamara, D. M. Cottrell, and J. Campos, "Image processing with the radial Hilbert transform: theory and experiments," Opt. Lett. **25**, 99-101 (2000).

Appendix A

Derivations

A.1 Phase Delay of a Tilted Plane Parallel Plate

The following is a derivation of the optical wavefront retardation introduced by tilting of a plane parallel glass plate in collimated light. Starting with Snell's Law [109] expressed for the angles of incidence, θ_i and refraction, θ_r relative to the surface normal $\hat{n}$ on entry to a glass plate of thickness t having refractive index n_r embedded in a medium of refractive index n_i:

$$n_i \sin\theta_i = n_r \sin\theta_r. \tag{A.1}$$

In this case, the rotation of the plate from normal incidence is equal to θ_i the angle of incidence (see Figure A.1).

If d is the physical path length within the plate then:

$$\sin\theta_r = \frac{\sqrt{d^2 - t^2}}{d}, \tag{A.2}$$

so that,

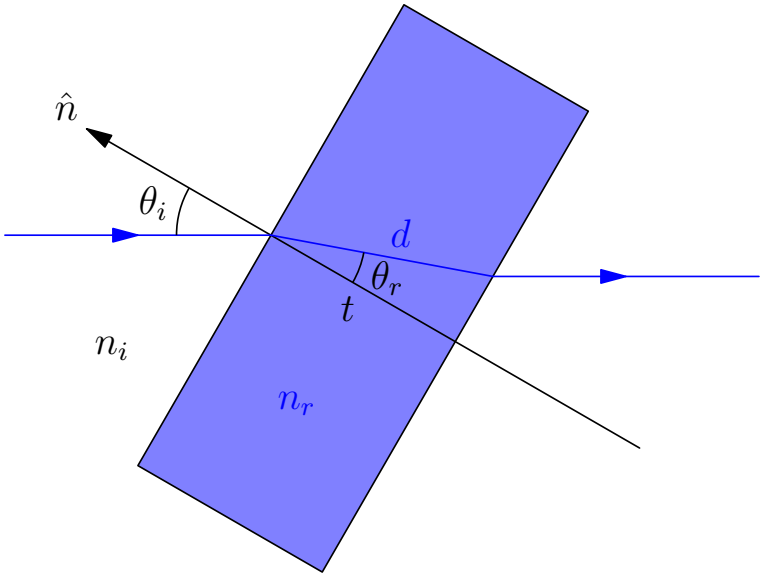

Figure A.1: Ray Deviation in a Tilted Plane Parallel Plate

$$\frac{n_i \sin\theta_i}{n_r} = \frac{\sqrt{d^2 - t^2}}{d}. \tag{A.3}$$

Solving for d as a function of θ_i leads to

$$d(\theta_i) = \frac{n_r t}{\sqrt{n_r^2 - n_i^2 \sin^2\theta_i}}. \tag{A.4}$$

The Optical Path Length (OPL), denoted Λ within the plate is the product of the refractive index of the plate and the physical path length as:

$$\Lambda(\theta_i) = n_r d = \frac{n_r^2 t}{\sqrt{n_r^2 - n_i^2 \sin^2\theta_i}}. \tag{A.5}$$

The Optical Path Difference (OPD), denoted $\Delta\Lambda$ between a ray passing through the tilted plate and untilted plate is

$$\Delta\Lambda(\theta_i) = n_r d - n_r t = \frac{n_r^2 t}{\sqrt{n_r^2 - n_i^2 \sin^2\theta_i}} - n_r t. \tag{A.6}$$

If instead of referencing the refractive indices to the speed of light *in vacuo*, they are referenced to that of the incident medium (typically air, having a refractive index very slightly above unity), then setting $n = {}^{n_r}/_{n_i}$ the OPD reduces to:

$$\Delta\Lambda(\theta_i) = \frac{n^2 t}{\sqrt{n^2 - \sin^2\theta_i}} - nt = nt\left(\frac{n}{\sqrt{n^2 - \sin^2\theta_i}} - 1\right). \tag{A.7}$$

If the OPD is to be expressed as a phase retardation $\Delta\phi$ in radians then the OPD is divided by the optical wavelength and multiplied by 2π as:

$$\Delta\phi(\theta_i) = \frac{2\pi\Delta\Lambda(\theta_i)}{\lambda} = \frac{2\pi nt}{\lambda}\left(\frac{n}{\sqrt{n^2 - \sin^2\theta_i}} - 1\right). \tag{A.8}$$

The wavelength, λ is measured in the reference medium.

Appendix B

Article in Press and Acceptance Correspondence

The email correspondence from the journal Current Applied Physics, as well as the Article in Press are provided below.

From: Naven Chetty <ChettyN3@ukzn.ac.za>
To: "Derek Griffith (dgriffith@csir.co.za)" <dgriffith@csir.co.za>
Date: 27/03/2015 09:12
Subject: FW: Article tracking [CAP_3911] - Accepted manuscript available online (unedited version)

-----Original Message-----
From: Elsevier - Article Status [mailto:Article_Status@elsevier.com]
Sent: 22 March 2015 09:22 PM
To: Naven Chetty
Subject: Article tracking [CAP_3911] - Accepted manuscript available online (unedited version)

Article title: Zernike-basis expansion of the fractional and radial Hilbert phase masks
Reference:: CAP3911
Journal title: Current Applied Physics
Corresponding author: Dr. N Chetty
First author: Dr. N Chetty
Accepted manuscript (unedited version) available online: 21-MAR-2015 DOI information: 10.1016/j.cap.2015.03.017

Dear Dr. Chetty,

We are pleased to inform you that your accepted manuscript (unformatted and unedited PDF) is now available online at:

http://authors.elsevier.com/sd/article/S1567173915000942

You might like to bookmark this permanent URL to your article.Please note access to the full text of this article will depend on your personal or institutional entitlements.
This version of your article has already been made available at this early stage to provide the fastest access to your article. It is not intended to be the final version of your article. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note changes to the article should not be requested at this stage.

Your article can already be cited using the year of online availability and the DOI as follows: Author(s), Article Title, Journal (Year), DOI.

Once the full bibliographic details (including volume and page numbering) for citation purposes are available, you will be alerted by e-mail.

To track the status of your article throughout the publication process, please use our article tracking service:

http://authors.elsevier.com/TrackPaper.html?trk_article=CAP3911&trk_surname=Chetty

Yours sincerely,
Elsevier Author Support

Current Applied Physics xxx (2015) 1–9

Contents lists available at ScienceDirect

Current Applied Physics

journal homepage: www.elsevier.com/locate/cap

Zernike-basis expansion of the fractional and radial Hilbert phase masks

N. Chetty*, D.J. Griffith

School of Chemistry and Physics, University of KwaZulu-Natal (PMB), Private Bag X01, Scottsville 3209, South Africa

ARTICLE INFO

Article history:
Received 7 January 2015
Accepted 20 March 2015
Available online xxx

ABSTRACT

The linear Hilbert phase mask or transform has found applications in image processing and spectroscopy. An optical version of the fractional Hilbert mask is considered here, comprising an imaging system with a circular, unobscured pupil in which a variable phase delay is introduced into one half of the pupil, split bilaterally.

The radial Hilbert phase mask is also used in image processing and to produce optical vortices which have applications in optical tweezers and the detection of exoplanets.

We subjected the fractional and radial Hilbert phase masks to Zernike function expansion in order to compute the image plane electromagnetic field distribution using Nijboer-Zernike theory. The Zernike functions form an orthogonal basis on the unit circle. The complex-valued Zernike expansion coefficients for these two phase masks were derived for use in the context of the Extended Nijboer-Zernike (ENZ) theory of image formation. The ENZ approach is of interest in that it allows a greater range of defocus to be dealt with, provides a simple means of taking a finite source size into account and has been adapted to high Numerical Aperture (NA) imaging applications.

Our image plane results for the fractional Hilbert mask were verified against a numerical model implemented in the commercial optical design and analysis code, Zemax®. It was found that the Nijboer-Zernike result converged to the Zemax® result from below as the number of Zernike terms in the expansion was increased.

1. Introduction

Many problems in optics relate to the diffraction propagation of electro-magnetic fields. In the modeling of imaging systems, computation of the shape of the wavefront emerging from the system exit pupil is often fairly straightforward and can be accomplished to adequate accuracy using simple geometrical ray-tracing. The final step, that of diffraction propagation of the emergent wavefront into the focal region is then the key computation in arriving at the complex field amplitude and hence the radiant intensity at or near focus. This involves evaluation of a diffraction integral, some of the most general being the Rayleigh-Sommerfeld integrals [1].

This technique of using geometrical optics to map the field in the exit pupil followed by numerical evaluation of a diffraction integral is one of the methods used in optical design and analysis software codes. The diffraction integral that is numerically evaluated can be selected to suit the problem. Various approximations can be used to simplify the most general diffraction integrals, including the Debye, Fresnel and Fraunhofer approximations [1]. The Fresnel and Fraunhofer approximations are of special significance in that diffraction propagation can be expressed in terms of linear and shift-invariant operations in frequency space giving rise to the domain of Fourier optics [2]. This permits application of all the analytical and numerical Fourier techniques to a broad range of optical diffraction problems.

Wavefront aberrations can be expanded in terms of the Zernike functions, which form an orthogonal set of basis functions on the unit circle. This covers the case of the unobscured, circular pupil which is very common in imaging systems. Not only does Zernike-basis expansion aid in the diagnostic interpretation of test interferograms, for example, it also allows for application of the Classical Nijboer-Zernike (CNZ) theory in optical analysis [1]. In CNZ theory, each Zernike function mode in the exit pupil gives rise to a proportional field contribution in the image plane [1].

* Corresponding author.
E-mail address: ChettyN3@ukzn.ac.za (N. Chetty).

http://dx.doi.org/10.1016/j.cap.2015.03.017
1567-1739/

More recently [3], the Extended Nijboer-Zernike (ENZ) theory has been introduced permitting coverage of a broader range of diffraction imaging phenomena. A detailed description of the ENZ approach with extensive references has been made available in [4]. Here, we apply the ENZ theory to analyse the circular pupil, fractional Hilbert mask [5] and perform a numerical verification of the result. ENZ expansion coefficients for the radial Hilbert phase mask on a circular pupil are also provided together with some illustrations of the Point Spread Function (PSF) in the focal region.

2. The Zernike functions

The normalized Zernike functions $\widehat{Z}_n^m(\rho)$ are composed of the product of a normalization factor N_n^m, a radial Zernike polynomial $R_n^m(\rho)$ and azimuthal sine or cosine factors for $n \geq 0$ and $m = -n,-n+2,\ldots,n-2,n$ as [6].

$$\widehat{Z}_n^m(\rho,\theta) = \begin{cases} N_n^m R_n^m(\rho)\cos m\theta & m \geq 0 \\ -N_n^m R_n^{|m|}(\rho)\sin m\theta & m < 0 \end{cases}. \tag{1}$$

The restriction on m to vary from $-n$ to n in increments of 2 implies that if n is odd, m is also odd and if n is even, m will be even or zero. Conversely, if m is odd then n is also odd and if m is even or zero then n is also even or zero.

Normalization of the Zernike functions is not always included in the definition (Equation (1)) by authors (excluded by Born and Wolf [1], included by Noll [7] and Thibos et al. [6] for example). If the normalization factors are excluded from the Zernike function definition, the normalization factors are then effectively incorporated into the expansion coefficients (denoted α_n^m in the following section).

While the product of the radial polynomial and cosine/sine factors always varies from −1 to +1 [8,9], the normalization factor N_n^m for a specific Zernike function is given by [6,7].

$$N_n^m = \sqrt{\frac{2(n+1)}{1+\delta_{m,0}}}, \tag{2}$$

where the Kronecker delta $\delta_{m,n}$ is

$$\delta_{m,n} = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}. \tag{3}$$

The radial Zernike polynomials can be written explicitly [1] as

$$R_n^m(\rho) = \sum_{s=0}^{p} \frac{(-1)^s (n-s)!}{s!(p-s)!(q-s)!}\rho^{n-2s} \tag{4}$$

where $p \equiv (n-|m|)/2$ and $q \equiv (n+|m|)/2$ are integers that are always zero or positive and the exclamation mark (!) denotes the factorial of the preceding expression. The radial polynomial coefficients are all integers which always sum to unity (i.e. $R_n^m(1) = 1$ for all m and n).

2.1. Zernike function orthogonality

The Zernike functions as defined in Equation (1) obey a set of inner product orthogonality relations [1]. The inner product of a function $\Phi(\rho,\theta)$ with the Zernike functions is defined as

$$\left\langle \Phi(\rho,\theta), \widehat{Z}_n^m(\rho,\theta) \right\rangle \equiv \frac{1}{\pi}\int_0^{2\pi}\int_0^1 \Phi(\rho,\theta)\widehat{Z}_n^m(\rho,\theta)\rho d\rho d\theta. \tag{5}$$

The inner products of the Zernike functions with themselves are

$$\left\langle \widehat{Z}_n^m(\rho,\theta), \widehat{Z}_{n'}^{m'}(\rho,\theta) \right\rangle = \delta_{m,m'}\delta_{n,n'}. \tag{6}$$

For the radial polynomials

$$\int_0^1 R_n^m(\rho)R_{n'}^m(\rho)\rho d\rho = \frac{\delta_{n,n'}}{2(n+1)}, \tag{7}$$

and

$$\int_0^{2\pi} \cos m\theta \cos m'\theta d\theta = \pi(1+\delta_{m,0})\delta_{m,m'}. \tag{8}$$

in the azimuthal direction. A similar relationship exists for the sine functions. Orthonormality implies that the integral of the product of any two Zernike functions over the unit circle is zero unless the two functions are identical (have the same n and m). In the latter case, the integral is unity.

Since the Zernike functions form a complete, orthogonal basis on the unit circle, any (pupil) function $\Phi(\rho,\theta)$ defined on the unit circle can be expanded in terms of the Zernike functions. The normalized Zernike coefficients $\widehat{\alpha}_n^m$ in the Zernike representation of a function are unique for a given pupil function, regardless of how many terms are fitted. It also means that the coefficients (normalized in the following instance) for reconstructing $\Phi(\rho,\theta)$ can be found using the inner product as

$$\widehat{\alpha}_n^m = \left\langle \Phi(\rho,\theta), \widehat{Z}_n^m(\rho,\theta) \right\rangle. \tag{9}$$

If the Zernike functions are not normalized by definition, the orthogonality relationship in Equation (6) will be

$$\left\langle Z_n^m(\rho,\theta), Z_{n'}^{m'}(\rho,\theta) \right\rangle = \frac{1+\delta_{m,0}}{2(n+1)}\delta_{m,m'}\delta_{n,n'} \tag{10}$$

$$= \frac{1}{N_n^m N_{n'}^{m'}}\delta_{m,m'}\delta_{n,n'}. \tag{11}$$

2.2. Properties and relationships of the Zernike radial polynomials

The Zernike functions and radial polynomials have been studied in detail [9,10] in the general context of orthogonal polynomials [11]. Some special values and identities of the Zernike radial polynomials that will be required are [10].

$$R_n^m(1) = 1, \tag{12}$$

$$R_n^m(0) = (-\delta_{m,0})^{n/2}, \tag{13}$$

$$\int_0^1 R_n^m(\rho)d\rho = \frac{(-1)^p}{n+1}, \tag{14}$$

$$\int_0^1 R_n^m(\rho)\rho d\rho = \begin{cases} \frac{1}{2} & m,n = 0 \\ \frac{(-1)^p m}{n(n+2)} & \text{otherwise} \end{cases}. \tag{15}$$

Equations (14) and (15) could not be found explicitly in existing

Please cite this article in press as: N. Chetty, D.J. Griffith, Zernike-basis expansion of the fractional and radial Hilbert phase masks, Current Applied Physics (2015), http://dx.doi.org/10.1016/j.cap.2015.03.017

literature and are therefore proven here. Starting with a Zernike polynomial relationship provided by Prata and Rusch [12] and having defined

$$S_n^m(s,\rho) \equiv \int R_n^m(\rho)\rho^s d\rho, \tag{16}$$

they provide

$$S_n^m(0,\rho) = \frac{1}{n+1}\left[\left(R_{n+1}^{m+1}(\rho) - R_{n+1}^{m+3}(\rho) + \ldots\right) - \left(R_{n-1}^{m+1}(\rho) - R_{n-1}^{m+3}(\rho) + \ldots\right)\right], \tag{17}$$

where the summations stop when the azimuthal (upper) index exceeds the radial index. This is now rewritten using summation notation as

$$S_n^m(0,\rho) = \frac{1}{n+1}\sum_{k=m+1,m+3\ldots}^{n+1}(-1)^{\frac{k-m-1}{2}}R_{n+1}^k(\rho) - \frac{1}{n+1}\sum_{k=m+1,m+3\ldots}^{n-1}(-1)^{\frac{k-m-1}{2}}R_{n-1}^k(\rho). \tag{18}$$

From Equation (16), the definite integral

$$\int_0^1 R_n^m(\rho)d\rho = S_n^m(0,\rho) - S_n^m(0,0) = S_n^m(0,1), \tag{19}$$

since from Equation (18) with Equation (13) it is clear that $S_n^m(0,0) = 0$. Then substituting ρ=1 in Equation (18) and using the identity in Equation (12) yields

$$S_n^m(0,1) = \frac{1}{n+1}\sum_{k=m+1,m+3\ldots}^{n+1}(-1)^{\frac{k-m-1}{2}} - \frac{1}{n+1}\sum_{k=m+1,m+3\ldots}^{n-1}(-1)^{\frac{k-m-1}{2}}. \tag{20}$$

The summations cancel term for term, except for the final term with $k = n+1$ in the first summation, so that

$$S_n^m(0,1) = \frac{(-1)^{\frac{n-m}{2}}}{n+1} = \frac{(-1)^p}{n+1}, \tag{21}$$

thus proving Equation (14).

Prata and Rusch [12] show that

$$S_n^m(1,\rho) = \frac{1}{2(n+1)}\left[(n+m+2)S_{n+1}^{m+1}(0,\rho) + (n-m)S_{n-1}^{m+1}(0,\rho)\right]. \tag{22}$$

From Equation (16), the definite integral

$$\int_0^1 R_n^m(\rho)\rho d\rho = S_n^m(1,\rho) - S_n^m(1,0) = S_n^m(1,1). \tag{23}$$

By substituting ρ=0 in Equation (22) it is clear that $S_n^m(1,0) = 0$. Then substituting ρ=1 in Equation (22) and using Equation (21),

$$S_n^m(1,1) = \frac{1}{2(n+1)}\left[(n+m+2)\frac{(-1)^{\frac{n-m}{2}}}{n+2} + (n-m)\frac{(-1)^{\frac{n-m-2}{2}}}{n}\right] \tag{24}$$

$$= \frac{1}{2(n+1)}\left[(n+m+2)\frac{(-1)^p}{n+2} - (n-m)\frac{(-1)^p}{n}\right] \tag{25}$$

$$= \frac{(-1)^p}{2(n+1)}\left[\frac{n+m+2}{n+2} - \frac{n-m}{n}\right] \tag{26}$$

$$= \frac{(-1)^p m}{n(n+2)}. \tag{27}$$

The final step is some straightforward algebra, thus proving Equation (15), except for the case $m = n = 0$, follows easily from. $R_0^0(\rho) = 1$.

3. Fractional Hilbert mask expansion

The goal here is to find the ENZ coefficients of the one-dimensional, fractional Hilbert mask on a circular, unobscured pupil. Optically, this is equivalent to splitting the pupil bilaterally and introducing a variable phase shift into one half.

The Zernike coefficients are computed using the inner product as given in Equation (9) and the inner product definition in Equation (5). The ENZ approach differs with reference to the classical Nijboer-Zernike (CNZ) approach in that the complex-valued EM pupil field amplitude is expanded instead of the real-valued wavefront error. In the ENZ case, the formalism is essentially the same as for the classical approach, except that the Zernike coefficients are now generally also complex-valued [4] and are usually denoted β_n^m. We will use A_n^m and B_n^m for the real and complex-valued Zernike coefficients for the circular pupil fractional Hilbert mask.

For a general, complex-valued pupil function $P(\rho,\theta)$ the normalized ENZ expansion is (using the inner product notation defined in Equation (5))

$$\widehat{\beta}_n^m = \left\langle P(\rho,\theta), \widehat{Z}_n^m(\rho,\theta)\right\rangle. \tag{28}$$

Note that in what follows, the ENZ expansions will only consider cosine terms ($m \geq 0$), and therefore any modulus signs around m will be dropped.

To perform the ENZ expansion of the circular fractional Hilbert mask, the circular pupil is first segmented bilaterally by expressing the (real-valued) pupil transmission $A(\theta)$ in the azimuthal coordinate, θ, as an infinite Fourier cosine series representing a square wave [13–15] as

$$A(\theta) = \frac{1}{2} + \frac{2}{\pi}\sum_{k=1}^{\infty}\frac{\sin\frac{k\pi}{2}}{k}\cos k\theta \tag{29}$$

$$= \frac{1}{2} + \frac{2}{\pi}\sum_{k=1,3,5\ldots}^{\infty}\frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta. \tag{30}$$

A Fourier series of this kind, when truncated, exhibits the Gibbs phenomenon (ripple and overshoot near discontinuities [14]) and requires many terms in the summation for reasonable convergence. This aspect will discussed in more detail in §III A.

The real transmission function $A(\theta)$ is ideally unity in the right half-plane (x>0) and zero in the left half-plane ($x < 0$). So for the

Please cite this article in press as: N. Chetty, D.J. Griffith, Zernike-basis expansion of the fractional and radial Hilbert phase masks, Current Applied Physics (2015), http://dx.doi.org/10.1016/j.cap.2015.03.017

Hilbert mask, the right half of the pupil is expressed in transmission as $A(\theta)$, and the left half of the pupil is then $1-A(\theta)$.

The full, complex-valued pupil with a phase retardation of $\Delta\phi$ in the right half is then written as

$$P(\rho,\theta) = A(\theta)e^{-i\Delta\phi} + 1 - A(\theta) = \left(e^{-i\Delta\phi} - 1\right)A(\theta) + 1. \tag{31}$$

This is illustrated in Fig. 1.

The real-valued transmission function $A(\theta)$ is subject to normalized (coefficients for normalized Zernike functions are denoted $\widehat{A}_n^m$, $\widehat{B}_n^m$, $\widehat{\alpha}_n^m$ etc.) Zernike expansion first as

$$\begin{aligned}\widehat{A}_n^m &= \left\langle A(\theta), \widehat{Z}_n^m(\rho,\theta)\right\rangle \\ &= \frac{1}{2}\left\langle \widehat{Z}_0^0, \widehat{Z}_n^m(\rho,\theta)\right\rangle + \frac{2}{\pi}\left\langle \sum_{k=1,3,5\ldots}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta, \widehat{Z}_n^m(\rho,\theta)\right\rangle \\ &= \frac{1}{2}\left\langle \widehat{Z}_0^0, \widehat{Z}_n^m(\rho,\theta)\right\rangle + \frac{2N_n^m}{\pi}\left\langle \sum_{k=1,3,5\ldots}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta, R_n^m(\rho)\cos m\theta\right\rangle.\end{aligned} \tag{32}$$

The final inner product in Equation (32) is evaluated using the definition in Equation (6) as

$$\begin{aligned}&\left\langle \sum_{k=1,3,5\ldots}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta, R_n^m(\rho)\cos m\theta\right\rangle \\ &= \frac{1}{\pi}\int_0^{2\pi}\int_0^1 \left[\sum_{k=1,3,5\ldots}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta\right] R_n^m(\rho)\cos m\theta\rho d\rho d\theta \\ &= \frac{1}{\pi}\int_0^{2\pi}\left[\sum_{k=1,3,5\ldots}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta\right]\cos m\theta d\theta \int_0^1 R_n^m(\rho)\rho d\rho \\ &= \frac{1}{\pi}\sum_{k=1,3,5\ldots}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k}\int_0^{2\pi}\cos k\theta\cos m\theta d\theta \int_0^1 R_n^m(\rho)\rho d\rho.\end{aligned} \tag{33}$$

Applying the cos-function orthogonality relationship given in Equation (8) as well as Equation (15) we notice that all terms in the summation vanish except for when $k = m$. Thus

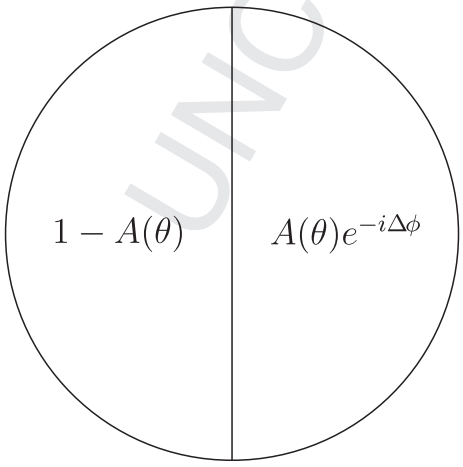

Fig. 1. Fractional Hilbert mask pupil function.

$$\begin{aligned}&\left\langle \sum_{k=1,3,5\ldots}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k}\cos k\theta, R_n^m(\rho)\cos m\theta\right\rangle \\ &= \frac{1}{\pi}\sum_{k=1,3,5\ldots}^{\infty} \frac{(-1)^{\frac{k-1}{2}}}{k}\pi\delta_{k,m}\int_0^1 R_n^m(\rho)\rho d\rho \\ &= \begin{cases} \dfrac{(-1)^{\frac{m-1}{2}}}{m}\dfrac{(-1)^{\frac{n-m}{2}}m}{n(n+2)} & m,n \text{ positive, odd} \\ 0 & \text{otherwise}\end{cases} \\ &= \begin{cases} \dfrac{(-1)^{\frac{n-1}{2}}}{n(n+2)} & m,n \text{ positive, odd} \\ 0 & \text{otherwise.}\end{cases}\end{aligned} \tag{34}$$

Substituting Equation (34) in Equation (32) gives

$$\widehat{A}_n^m = \begin{cases} \dfrac{1}{2} & m = n = 0 \\ \dfrac{2N_n^m(-1)^{\frac{n-1}{2}}}{\pi n(n+2)} & m,n \text{ odd} \\ 0 & m,n \text{ even.}\end{cases} \tag{35}$$

It is noted that these coefficients are independent of m (the coefficients are the same across all values of m for each n).

The fractional Hilbert mask pupil function $P(\rho,\theta)$ is then expanded using the ENZ procedure as

$$\begin{aligned}\widehat{B}_n^m &= \left\langle P(\rho,\theta), \widehat{Z}_n^m(\rho,\theta)\right\rangle \\ &= \left\langle (e^{-i\Delta\phi} - 1)A(\theta) + 1, \widehat{Z}_n^m(\rho,\theta)\right\rangle \\ &= (e^{-i\Delta\phi} - 1)\widehat{A}_n^m + \left\langle \widehat{Z}_0^0, \widehat{Z}_n^m(\rho,\theta)\right\rangle.\end{aligned} \tag{36}$$

The $\widehat{B}_n^m$ coefficients can thus be written down from the $\widehat{A}_n^m$ coefficients (Equation (35)) as

$$\widehat{B}_n^m = \begin{cases} \dfrac{(e^{-i\Delta\phi} + 1)}{2} & m = n = 0 \\ \dfrac{2N_n^m\left(e^{-i\Delta\phi} - 1\right)(-1)^{\frac{n-1}{2}}}{\pi n(n+2)} & m,n \text{ odd} \\ 0 & m,n \text{ even.}\end{cases} \tag{37}$$

This solution for $\widehat{B}_n^m$ has the required property that if $\Delta\phi \to 0$ then $\widehat{B}_n^m \to 0$ for all positive m and n while $\widehat{B}_0^0 \to 1$. These coefficients are also independent of m.

3.1. Gibbs phenomenon

Formulation of the fractional Hilbert mask pupil function involved the use of a square wave, synthesized as a Fourier cosine sequence $A(\theta)$ given in Equation (29). This series has relatively poor convergence and exhibits ripple (or ringing) and overshoot artifacts near the discontinuities, known as "Gibbs Phenomenon" [14] when the series is truncated in numerical work. It follows that these artifacts will be propagated by the numerical ENZ process to the focal region results when the series is necessarily truncated.

The Gibbs phenomenon can be suppressed to a significant extent (but not eliminated) using a filtering factor called the

Please cite this article in press as: N. Chetty, D.J. Griffith, Zernike-basis expansion of the fractional and radial Hilbert phase masks, Current Applied Physics (2015), http://dx.doi.org/10.1016/j.cap.2015.03.017

Lanczos σ-factor [14]. If the series is to be truncated at $n = N$, the Fourier coefficients are multiplied by $\mathrm{sinc}\frac{n}{N}\infty$ where $\mathrm{sinc}(x)\equiv\sin\pi x/\pi x$ and the summation is performed up to $n = N-1$. So truncating the series at $n = N$ and including the Lanczos σ-factor, the $\widehat{B}_n^m$ coefficients for normalized Zernike functions become

$$\widehat{B}_n^m = \begin{cases} \dfrac{e^{-i\Delta\phi}+1}{2} & m=n=0 \\ \dfrac{2N_n^m\left(e^{-i\Delta\phi}-1\right)(-1)^{\frac{n-1}{2}}\mathrm{sin\,c}\dfrac{n}{N}}{\pi n(n+2)} & m,n \text{ odd}, n<N \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

The corresponding result for the $\widehat{A}_n^m$, being the normalized Zernike coefficients for $A(\theta)$ follows from Equation (35) as

$$\widehat{A}_n^m = \begin{cases} \dfrac{1}{2} & m=n=0 \\ \dfrac{2N_n^m(-1)^{\frac{n-1}{2}}\mathrm{sin\,c}\dfrac{n}{N}}{\pi n(n+2)} & m,n \text{ odd}, n<N \\ 0 & \text{otherwise.} \end{cases} \tag{39}$$

As an illustration, $A(\theta)$ is plotted for a summation of 50 terms with and without the Lanczos σ-factor in Fig. 2. Clearly the inclusion of the σ-factor helps to suppress the Gibbs phenomenon. The first discontinuity is at $\theta = \pi/2$.

With or without the Lanczos σ-factor for suppression of the Gibbs phenomenon, it is necessary to compute the Zernike radial polynomials to high order for numerical verification purposes. The coefficients (see Equation (4)) are ratios of factorials which become large, computationally expensive and potentially inaccurate from $n = 40$ onward [16]. Alternative techniques for computation of the polynomial coefficients are then required. Janssen and Dirksen [16] and Vlcek and Sovka [17] have provided solutions which exploit the relationship of the Zernike radial polynomials to Chebyshev polynomials and the more general Jacobi polynomials.

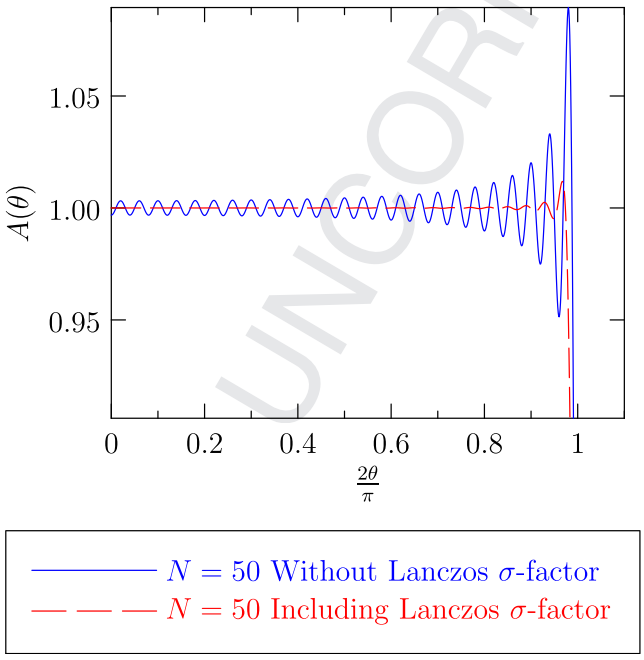

Fig. 2. $A(\theta)$ truncated at 50 terms.

4. ENZ propagation from pupil to focal region

The essence of the ENZ approach [3] is that each Zernike term in the complex-valued pupil expansion corresponds to a field contribution in the focal region. The complex-valued, scalar field amplitude, U, in the focal region is written as

$$U\left(\rho',\theta';d_f\right) = \sum_{m,n}\beta_n^m U_n^m\left(\rho',\theta';d_f\right), \tag{40}$$

where $\beta_n^m U_n^m$ is the contribution from $\beta_n^m Z_n^m$ in the exit pupil, the lateral coordinates in the focal region are (ρ',θ'), with a relative defocus from the best focal plane in the axial direction of d_f. Janssen [3] evaluated the focal region contributions U_n^m for substantial defocus d_f (relative to CNZ), by deriving functions $V_n^m(\rho',d_f)$ such that

$$U_n^m\left(\rho',\theta';d_f\right) = 2i^m V_n^m\left(\rho',d_f\right)\cos m\theta', \tag{41}$$

where the radial and defocus function V_n^m is an integral over Zernike polynomial, radially-weighted Bessel functions, J_m of the first kind and order m,

$$V_n^m\left(\rho',d_f\right) = \int_0^1 \exp\left\{id_f\rho^2\right\}R_n^m(\rho)J_m(2\pi\rho\rho')\rho d\rho. \tag{42}$$

Note that the focal region coordinates and defocus parameter d_f are not in absolute length units here, but expressed relative to the Numerical Aperture (NA) of imaging and the wavelength of light under consideration. Exit pupil and focal region coordinates are illustrated in §VI A.

Power-Bessel series expansion on V_n^m produces [3].

$$V_n^m\left(\rho',d_f\right) = e^{id_f}\sum_{l=1}^{\infty}\left(-2id_f\right)^{l-1}\sum_{j=0}^{p}v_{lj}\frac{J_{m+l+2j}(2\pi\rho')}{l(2\pi\rho')^l}, \tag{43}$$

in which $p = (n-m)/2$ (as before for positive m). The v_{lj} factor is computed as

$$v_{lj} = (-1)^p(m+l+2j)\binom{m+j+l-1}{l-1}\binom{j+l-1}{l-1} \times\binom{l-1}{p-j}\Big/\binom{q+l+j}{l}, \tag{44}$$

with $p = (n+m)/2$ (as before for positive m).

For defocus parameter $d_f = 0$, the result of the Classical Nijboer-Zernike (CNZ) theory can be used in which the summation over l in Equation (43) is reduced to a single Bessel term as [4].

$$V_n^m(\rho',0) = \int_0^1 R_n^m(\rho)J_m(2\pi\rho\rho')\rho d\rho = (-1)^{\frac{n-m}{2}}\frac{J_{n+1}(2\pi\rho')}{2\pi\rho'}. \tag{45}$$

The CNZ result in Equation (45) is useful for consistency checking and for the initial evaluation of convergence. Also, the evaluation of the v_{lj} binomial factors as given in Equation (44) can

Please cite this article in press as: N. Chetty, D.J. Griffith, Zernike-basis expansion of the fractional and radial Hilbert phase masks, Current Applied Physics (2015), http://dx.doi.org/10.1016/j.cap.2015.03.017

become numerically problematic at high order.

4.1. Pupil and focal region coordinates

The radial coordinate ρ in the pupil is normalized relative to the physical pupil radius. The Cartesian coordinates (x', y', d_f) and polar coordinates (ρ', θ') in the focal region are also normalized relative to absolute coordinates (x,y,z) as follows:

$$\begin{aligned} x' &= x\frac{NA}{\lambda} = \rho' \cos\theta' \\ y' &= y\frac{NA}{\lambda} = \rho' \sin\theta' \\ d_f &= 2\frac{\pi}{\lambda} z\left(1 - \sqrt{1 - NA^2}\right), \end{aligned} \tag{46}$$

where NA is the Numerical Aperture (NA) of imaging and λ is the light wavelength.

The relative coordinate scheme is illustrated in Fig. 3. Here the system exit pupil P with absolute outer edge radius R and relative radial coordinate ρ, has a spherical wavefront emerging from the exit pupil and converging on the image plane origin at O.

4.2. Zernike function normalization in ENZ practice

Braat et al. [4] exclude the normalisation factor N_n^m in their definition of the Zernike functions. In that case, the expressions for the unnormalized coefficients A_n^m and B_n^m for the fractional Hilbert mask (with reference to Equations (39) and (38), including the Lanczos σ-factor) will include the square of the N_n^m factors and

$$A_n^m = \begin{cases} \dfrac{1}{2} & m = n = 0 \\ \dfrac{2\left[N_n^m\right]^2 (-1)^{\frac{n-1}{2}} \operatorname{sinc}\dfrac{n}{N}}{\pi n(n+2)} & m, n \text{ odd}, n < N \\ 0 & \text{otherwise}, \end{cases} \tag{47}$$

with

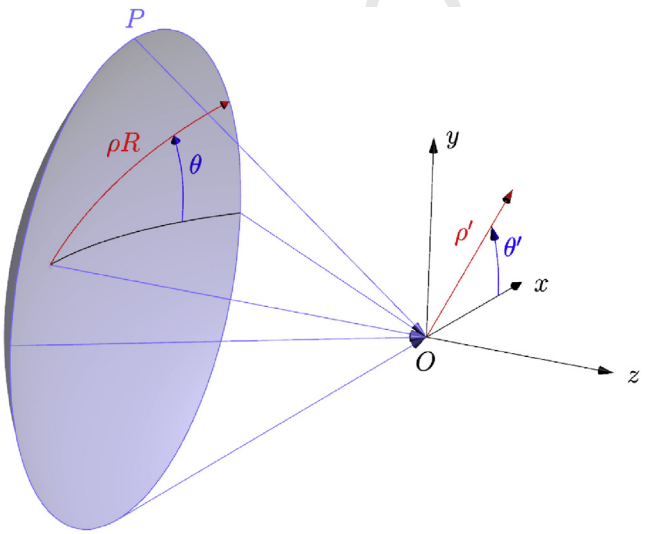

Fig. 3. Exit pupil (P) and focal region (O) coordinates.

$$B_n^m = \begin{cases} \dfrac{e^{-i\Delta\phi} + 1}{2} & m = n = 0 \\ \dfrac{2\left[N_n^m\right]^2 \left(e^{-i\Delta\phi} - 1\right)(-1)^{\frac{n-1}{2}} \operatorname{sinc}\dfrac{n}{N}}{\pi n(n+2)} & m, n \text{ odd}, n < N \\ 0 & \text{otherwise}. \end{cases} \tag{48}$$

4.3. ENZ scope and limitations

The scope and limitations of the application of ENZ tools *circa* 2010 is covered in detail by Van Haver [18]. The accuracy of this ENZ result relative to more accurate (numerical) Rayleigh integral computations is expected to degrade in both the low and high NA regimes. For optical wavelengths on the order of 550 nm in the visible spectrum, the applicable range is taken to be $0.05 < NA < 0.6$. Other limitations relating to vector and associated polarization effects are assumed negligible in the current context.

5. Numerical verification

5.1. CNZ convergence

The convergence of the CNZ result for the circular pupil, fractional Hilbert mask (Equations (40), (41) and (45) with complex-valued coefficients B_n^m computed with Equation (48) or 38) was tested with increasing N (the maximum value of n). A cross-section of the PSF intensity (irradiance) along the x-axis was computed for a sequence of increasing N with $\Delta\phi = \pi, NA = 0.1$ and $\lambda = 0.5\ \mu m$. The CNZ result without the normalization factor is plotted in Fig. 4. This result is shown relative to the Non-Sequential Component (NSC) model result from the optical analysis package Zemax®. The Zemax® result will be discussed in more detail in §VI.

A horizontal line at a relative irradiance which is 4 orders of magnitude below peak has been plotted in Fig. 4. This is assumed to

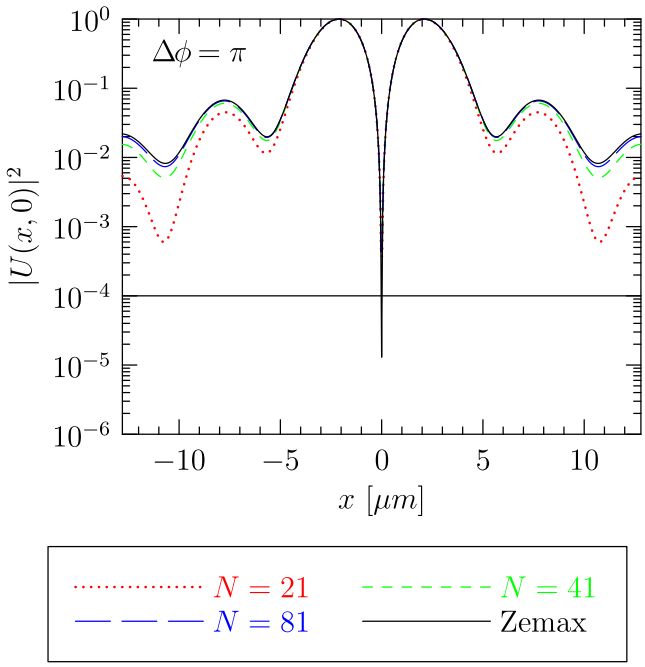

Fig. 4. Circular Fractional Hilbert Pupil Mask, $\log_{10}$ Normalized Image PSF x-Cross Section (CNZ), $\Delta\phi = \pi$, $NA = 0.1$, $\lambda = 0.5\ \mu m$, $d_f = 0$.

Please cite this article in press as: N. Chetty, D.J. Griffith, Zernike-basis expansion of the fractional and radial Hilbert phase masks, Current Applied Physics (2015), http://dx.doi.org/10.1016/j.cap.2015.03.017

be measurable with high dynamic range Focal Plane Array (FPA) detectors.

The x-cross section convergence of the CNZ result relative to NSC Zemax® results for other values of $\Delta\phi$ are plotted in Figs. 5–7.

For $\Delta\phi = 0$ (Fig. 5), there is only a single CNZ term, providing the normal incoherent diffraction-limited PSF for an unobscured, circular pupil (the Airy disk).

For non-zero $\Delta\phi$ convergence appears to be from below. The image-plane PSF for a number of different values of $\Delta\phi$ was computed using the CNZ result (Equation (45), d_f=0) and the $\log_{10}$ of the normalized irradiance (square modulus of the complex-valued scalar field) is plotted in Fig. 8 for $NA = 0.1$, $\lambda = 0.5\,\mu m$ and $N = 41$.

The result for $\Delta\phi$=0 in Fig. 8 yields the expected diffraction-limited PSF for $NA = 0.1$, $\lambda = 0.5\,\mu m$, corresponding to $m = n = 0$ in Equation (45). The full, central dark interference null is only formed close to $\Delta\phi = \pi$.

Braat et al. [4] provide Matlab® code for evaluation of the V_n^m functions, comprising implementation of Equation (43). This code was used for the computation of the field in the focal plane for the circular pupil, fractional Hilbert mask. The result was verified to be the same for the $d_f = 0$ case as for the CNZ result.

In the process it was noted that the radial coordinate input to the Matlab® function for V_n^m must be scaled by 2π in addition to normalizing as per Equation (46).

6. Comparison to Zemax® results

Zemax® is a multipurpose, optical design and analysis software code. Two methods of modeling the circular, fractional Hilbert mask problem were attempted. In the first method, the phase piston ($\Delta\phi$) was introduced into the optical layout using a Zemax® surface type called the "Zernike Standard Phase" surface. This surface directly introduces a phase advance/retardation into the wavefront through specification of the Zernike-basis coefficients. For various reasons (more detail is given in §VI A), this method did not work well and was useful only to help verify the

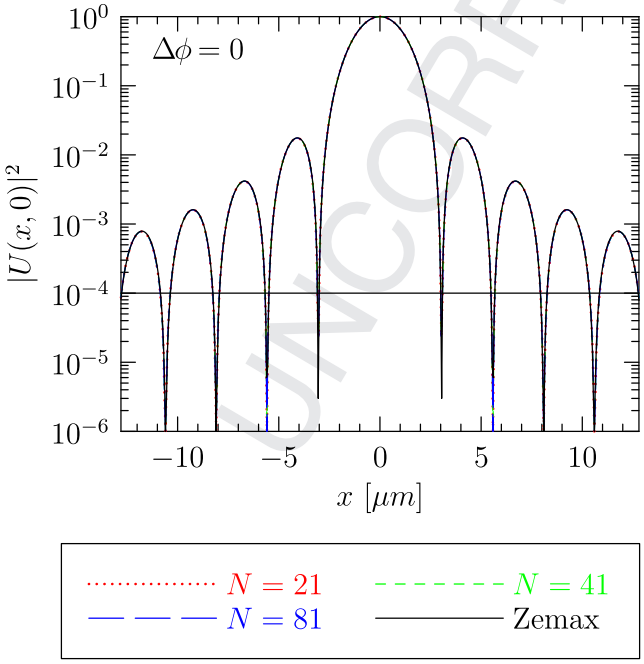

Fig. 5. Circular Fractional Hilbert Pupil Mask, $\log_{10}$ Normalized Image PSF x-Cross Section (CNZ), $\Delta\phi = 0$, $NA = 0.1$, $\lambda = 0.5\ \mu m$, $d_f = 0$.

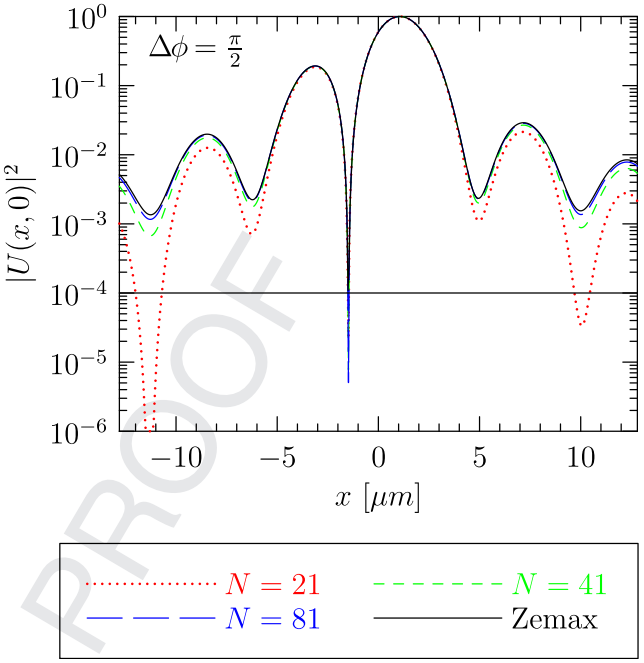

Fig. 6. Circular Fractional Hilbert Pupil Mask, $\log_{10}$ Normalized Image PSF x-Cross Section (CNZ).$\Delta\phi = \frac{\pi}{2}$, $NA = 0.1$, $\lambda = 0.5\mu m$, $d_f = 0$.

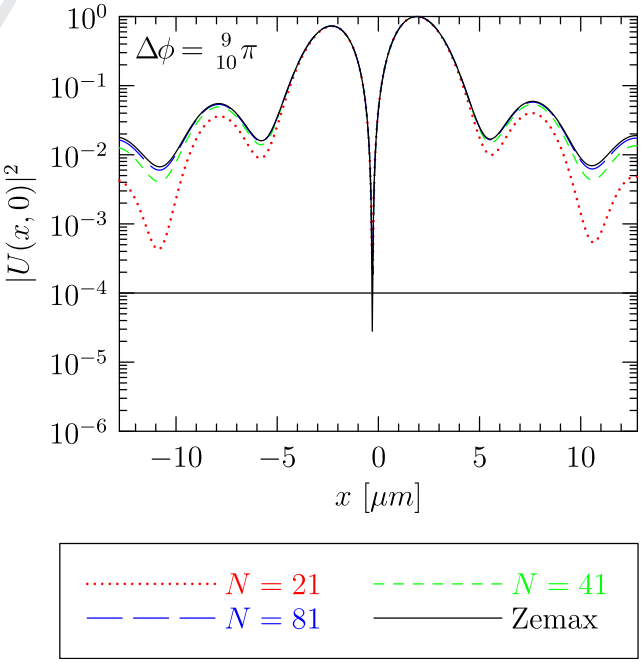

Fig. 7. Circular Fractional Hilbert Pupil Mask, $\log_{10}$ Normalized Image PSF x-Cross Section (CNZ).$\Delta\phi = \frac{9\pi}{10}$, $NA = 0.1$, $\lambda = 0.5\mu m$, $d_f = 0$.

Zernike coefficients $\widehat{A}_n^m$ for normalized Zernike functions (Equation (39)).

The second Zemax® modeling method entailed introduction of the phase piston $\Delta\phi$ using a Non-Sequential Component (NSC) model of a tilted glass plate phase retarder as illustrated in Fig. 9. It was necessary to use two pairs of plates in sequence, with one sequential pair tilted in opposite directions in order to undo the small lateral beam displacement caused by the first tilted plate.

Please cite this article in press as: N. Chetty, D.J. Griffith, Zernike-basis expansion of the fractional and radial Hilbert phase masks, Current Applied Physics (2015), http://dx.doi.org/10.1016/j.cap.2015.03.017

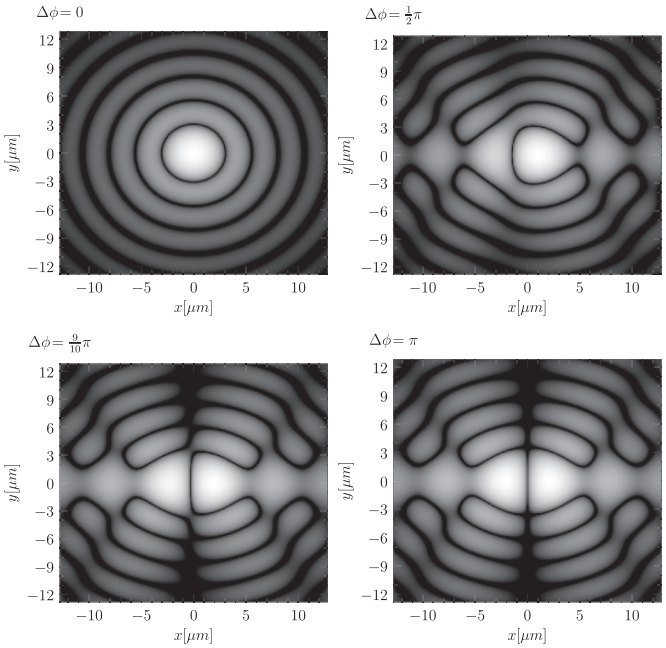

Fig. 8. Circular Fractional Hilbert Pupil Mask, log Normalized Image PSF Irradiance (CNZ), $NA = 0.1$, $\lambda = 0.5$ μm, $d_f = 0$, $N = 41$.

Once the Zemax® NSC model of the plate retarder had been set up and verified, the cross-section of the PSF was computed, using direct summation of Huygens wavelets. This is thought to be the more accurate (and definitely more time-consuming) method with respect to the more routine FFT approach. Besides the Huygens wavelet summation and FFT methods, Zemax® also offers a Physical Optical Propagation (POP) method of analysis. The Zemax® result from the Huygens wavelet-summation method is shown as a reference in Figs. 4–7. The CNZ result does converge to the Zemax result, but quite a large number of terms ($N > 80$) are required for close agreement.

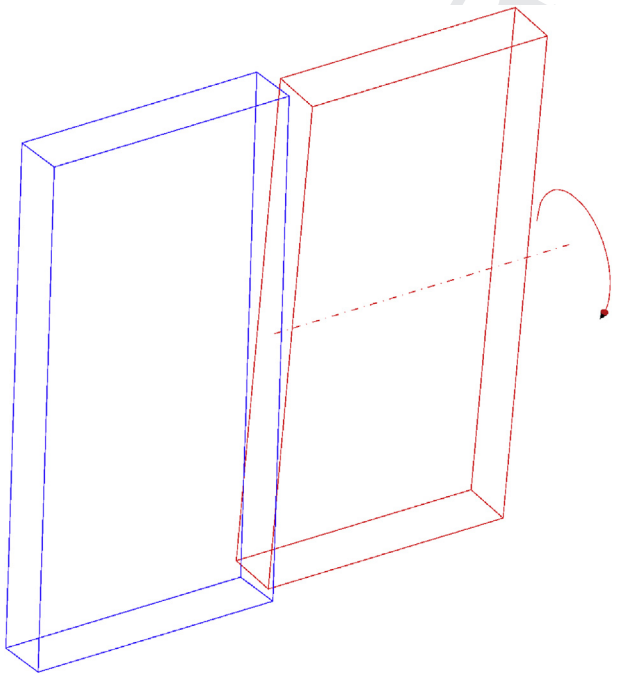

Fig. 9. Tilted plane parallel glass plate phase retarder as fractional Hilbert mask implementation. The phase retarder would be placed in the system pupil.

6.1. Zernike functions in Zemax

Zemax® follows the Noll [7] indexing scheme for the Zernike ("standard" as opposed to "fringe") functions which maps $Z_n^m \rightarrow Z_j$ according to sequence A176988 in the On-Line Encyclopedia of Integer Sequences. This mapping appears to lack the virtue of a simple formula, but follows the rule that odd j are assigned to values of $m < 0$ (and even j to all $m > 0$) and smaller values of j are assigned to smaller values of $|m|$.

Zemax only allows up to $j = 231$ Noll terms in the Zernike series. This corresponds to all terms up to $n = 21$. This is not really sufficient to allow for an accurate computation of the fractional Hilbert pupil mask situation using the Zernike Standard Phase surface in Zemax. Zemax® also follows Noll [7] in that the Zernike functions include the normalization factor.

6.2. Computation of circular pupil, fractional Hilbert mask PSF using Zemax®

The circular pupil, fractional Hilbert mask was modeled in Zemax® using a Non-Sequential Component (NSC) group and implemented as a tilted plate phase retarder illustrated in Fig. 9. Zemax® currently offers three different methods of computing the PSF, namely the direct summation of Huygens wavelets, an FFT technique and a Physical Optics Propagation (POP) algorithm. All three methods produced essentially the same results illustrated in Figures 4 to 8.

The Huygens PSF results from Zemax® are visually indistinguishable from the results shown in Fig. 8. Note that the x-axis cross sections of the Zemax® Huygens PSF image results are plotted as the reference curves in Figs. 4–7.

7. The radial Hilbert phase mask

The radial Hilbert phase mask is equivalent to a phase vortex used in [19] and in vortex coronagraph concepts for exoplanet detection [20]. In this case the pupil phase is written as

$$P(\theta) = e^{il\theta}, \tag{49}$$

where the integer vortex topological charge is $l \geq 1$. We performed unnormalised Zernike-basis expansion using the ENZ approach and for $m \geq 0$ as

$$\begin{aligned}
\beta_n^m &= \left\langle e^{il\theta}, Z_n^m(\rho,\theta) \right\rangle \\
&= \frac{1}{\pi}\int_0^{2\pi}\int_0^1 e^{il\theta} Z_n^m(\rho,\theta)\rho d\rho d\theta \\
&= \frac{1}{\pi}\int_0^{2\pi}\int_0^1 e^{il\theta} R_n^m(\rho)\cos m\theta \rho d\rho d\theta \\
&= \frac{1}{\pi}\int_0^{2\pi} e^{il\theta}\cos m\theta d\theta \int_0^1 R_n^m(\rho)\rho d\rho \\
&= \frac{1}{\pi}\int_0^{2\pi} (\cos l\theta + i\sin l\theta)\cos m\theta d\theta \int_0^1 R_n^m(\rho)\rho d\rho \\
&= \frac{\delta_{l,m}(-1)^p m}{n(n+2)}.
\end{aligned} \tag{50}$$

Please cite this article in press as: N. Chetty, D.J. Griffith, Zernike-basis expansion of the fractional and radial Hilbert phase masks, Current Applied Physics (2015), http://dx.doi.org/10.1016/j.cap.2015.03.017

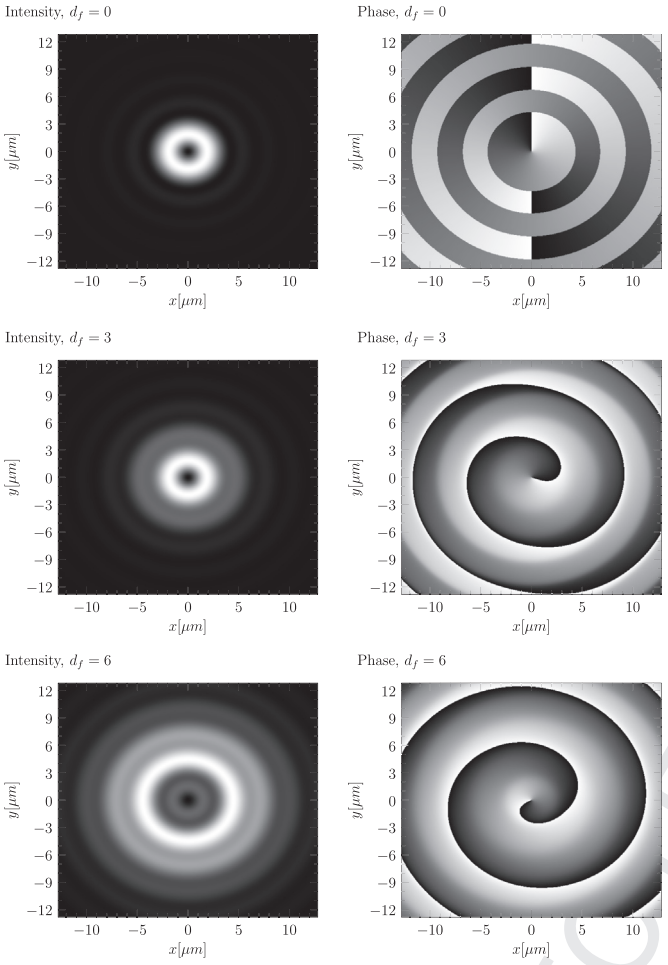

Fig. 10. Optical vortex focal region normalized linear intensity and phase for charge l = 1. Normalised intensity varies from 0 (black) to 1 (white) and phase varies from 0 (black) to 2π(white), $NA = 0.1$, $\lambda = 0.5$ μm.

The final step arises from the orthogonality relations of the sin and cos functions and application of Equation (15). For $m < 0$, a similar argument yields

$$\beta_n^m = \frac{-i\delta_{l,|m|}(-1)^p|m|}{n(n+2)} = -i\beta_n^{-m}. \tag{51}$$

The vortex charge l selects the corresponding azimuthal orders $|m| = l$ only, but otherwise there are non-zero coefficients for all n. The ENZ focal region reconstruction proceeds according to Equation (40). Fig. 10 shows the intensity and phase in focus ($d_f = 0$) and out of focus ($d_f = 3$ and $d_f = 6$).

8. Conclusion

The Zernike functions were reviewed as an orthogonal set of basis functions on the unit circle. The Zernike function basis has found numerous applications in optics and other disciplines such as image processing. A selection of Zernike function and Zernike radial polynomial relations were presented in the context of the Classical Nijboer-Zernike (CNZ) theory of optical aberrations. The more recent, and growing, Extended Nijboer-Zernike (ENZ) approach [3,18] has expanded the range of optical diffraction problems that can be modeled and analyzed using an efficient semi-analytical procedure rooted in the use of Zernike-basis expansion.

The linear, fractional Hilbert mask has found applications in optics [5] and image processing [21]. The main result presented here was derivation of the Extended Nijboer-Zernike expansion coefficients for the circular pupil, fractional Hilbert mask. The convergence of the resulting infinite series, when truncated, was improved by introduction of the Lanczos σ-factor [14]. The ENZ coefficients for the circular pupil, fractional Hilbert mask were verified using numerical techniques, especially that of a comparison to results from the Zemax® optical design and analysis code.

The circular pupil ENZ expansion coefficients for the radial Hilbert phase mask, which is identified with the optical vortex, have also been provided together with illustrations of the focal region PSF intensity and phase.

Our future work in this area will relate to the use of Hilbert masks to produce axial irradiance nulls for high dynamic range measurements of scattering phase functions and also for aberration retrieval using irradiance measurements in the focal region.

References

[1] M. Born, E. Wolf, Principles of Optics, seventh ed., Cambridge University Press, 1999.
[2] J.W. Goodman, Introduction to Fourier Optics, McGraw-Hill, 1968.
[3] A.J.E.M. Janssen, Extended Nijboer-Zernike approach for the computation of optical point-spread functions, J. Opt. Soc. Am. A 19 (5) (2002) 849–857.
[4] J.J.M. Braat, P. Dirksen, S. van Haver, A.J.E.M. Janssen, Detailed description of the ENZ approach. http://www.nijboerzernike.nl/, 2013.
[5] A.W. Lohmann, D. Mendlovic, Z. Zalevsky, Fractional Hilbert transform, Opt. Lett. 21 (1996) 281–283.
[6] L. Thibos, R.A. Applegate, J.T. Schwiegerling, R. Webb, Standards for reporting the optical aberrations of eyes, in: Vision Science and its Applications, Optical Society of America, 2000 p. SuC1.
[7] R.J. Noll, Zernike polynomials and atmospheric turbulence, J. Opt. Soc. Am. 66 (1976) 207–211.
[8] A.J. Janssen, New analytic results for the Zernike circle polynomials from a basic result in the Nijboer-Zernike diffraction theory, J. Eur. Opt. Soc. – Rapid Publ. 6 (2011).
[9] G. Szegö, Orthogonal polynomials, in: American Mathematical Society Colloquium Publications, vol. 23, American Mathematical Society, 1939.
[10] W.J. Tango, The circle polynomials of Zernike and their application in optics, Appl. Phys. 13 (1977) 327–332.
[11] M. Abramowitz, I.A. Stegun (Eds.), Handbook of Mathematical Functions, tenth ed.Applied Mathematics Series, vol. 55, NIST, 1964.
[12] J. Aluizio Prata, W.V.T. Rusch, Algorithm for computation of Zernike polynomials expansion coefficients, Appl. Opt. 28 (1989) 749–754.
[13] J.P. Boyd, Chebyshev and Fourier Spectral Methods, second ed., Dover Publications, 2001. http://www-personal.umich.edu/jpboyd/aaabook_9500may00.pdf.
[14] A. Jerri, The Gibbs Phenomenon in Fourier Analysis, Splines and Wavelet Approximations, in: Mathematics and its Applications, vol. 446, Springer, 1998.
[15] K. Howell, Principles of Fourier Analysis, Studies in Advanced Mathematics, Taylor & Francis, 2010.
[16] A. Janssen, P. Dirksen, Computing Zernike polynomials of arbitrary degree using the discrete fourier transform, J. Eur. Opt. Soc. - Rapid Publ. 2 (2007).
[17] M. Vlcek, P. Sovka, Zernike polynomials and their spectral representation, Proc. 2013 Int. Conf. Electron. Signal Process. Commun. Syst. (2013). Q3
[18] S. Van Haver, The Extended Nijboer-Zernike Diffraction Theory and its Applications (Ph.D. thesis), TU Delft, 2010, http://repository.tudelft.nl/view/ir/uuid:8d96ba75-24da-4e31-a750-1bc348155061/.
[19] D.G. Grier, A revolution in optical manipulation, Nature 424 (2003) 810816.
[20] G. Foo, D.M. Palacios, Grover A. Swartzlander, Optical vortex coronagraph, Opt. Lett. 30 (2005) 33083310.
[21] J.A. Davis, D.E. McNamara, D.M. Cottrell, J. Campos, Image processing with the radial Hilbert transform: theory and experiments, Opt. Lett. 25 (2000) 99–101.

Please cite this article in press as: N. Chetty, D.J. Griffith, Zernike-basis expansion of the fractional and radial Hilbert phase masks, Current Applied Physics (2015), http://dx.doi.org/10.1016/j.cap.2015.03.017