



**FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA**

**MODELLING AND SIMULATION OF LIGHT
SCATTERING IN OPTICAL SYSTEMS**

Dissertation

for the acquisition of the academic title

Doctor Rerum Naturalium (Dr. rer. nat.)

submitted to the Council of the Faculty of Physics and Astronomy of
Friedrich-Schiller-Universität Jena

By M.Sc. **Xiang Lu**

Born in Changyi, Shandong Province, China on 01.04.1991

Gutachter:

1. Prof. Dr. Herbert Gross, Friedrich-Schiller-Universität Jena
2. Prof. Dr. Alois Herkommer, Universität Stuttgart
3. Prof. Dr. Alexander Rohrbach, Albert-Ludwigs-Universität Freiburg

Date of the disputation: 28.10.2021

Zusammenfassung

Die Lichtstreuung weicht das Licht von den normalen Trajektorien im optischen System ab und kann zu einer Verschlechterung der optischen Leistung führen, hauptsächlich des Kontrasts. Daher ist die Simulation der Lichtstreuung in optischen Systemen für die Bewertung der Systemleistung von entscheidender Bedeutung. Die derzeitigen Simulationsmethoden der Lichtstreuung basieren jedoch hauptsächlich auf Monte Carlo Raytracing, was äußerst zeitaufwändig ist. Obwohl es Methoden gibt, um die strahlbasierte Simulation der Lichtstreuung zu beschleunigen, ist ihre Genauigkeit begrenzt, oder sie können für optische Systeme ohne Kreissymmetrie kaum angewendet werden. Darüber hinaus sind die strahlbasierten Methoden vollständig geometrisch und können nur die Streuung durch Oberflächenrauheit simulieren, während sie die Kleinwinkelstreuung durch mid-spatial-frequency (MSF) Fehlern nicht genau modellieren können, da die physikalischen Aspekte von Licht wie z. B. partiell Kohärenz, Beugung und Interferenz modelliert werden müssen.

In dieser Arbeit schlagen wir zunächst eine quasi-analytische Phasenraummethode zur Simulation der Lichtstreuung in optischen Systemen vor, die die Anregung und Akzeptanz der Lichtstreuung im Phasenraum direkt koppelt. Da das Phasenraummethode keine zufällige Abtastung der Strahlrichtung beinhaltet, ist es viel effizienter und liefert Ergebnisse mit einem hohen Signal-Rausch-Verhältnis (SNR). Wir implementieren zunächst die Phasenraummethode, um den Autofluoreszenzeffekt in Mikroskoplinsen zu simulieren, auf deren Grundlage eine umfassende Analyse des Autofluoreszenzeffekts verschiedener Arten von Mikroskoplinsen vorgestellt wird. Anschließend erweitern wir die Phasenraummethode, um die Oberflächenstreuung in optischen Systemen zu simulieren. Wir evaluieren das erweiterte Phasenraummodell in Spiegelsystemen mit und ohne Kreissymmetrie und schlagen drei Möglichkeiten zur Implementierung der Phasenraummethode vor, die für verschiedene Arten von optischen Systemen geeignet sind. Durch Vergleichen der drei Implementierungen werden die Anwendbarkeit, Effizienz, Genauigkeit und Einschränkungen der Phasenraummethode diskutiert.

Um die Ausbreitung von teilweise kohärentem Licht in trüben Medien oder durch statistisch gestörte Oberflächen zu simulieren, schlagen wir außerdem eine auf der Wigner-Funktion basierende Methode vor, bei der die partielle Kohärenz von Licht genau modelliert wird, und die Streuung des teilweise kohärenten Lichts durch mehrere Streuflächen effizient simuliert werden kann. Da die Wigner-Funktion von Licht gleichzeitig die Lichtverteilung im Winkel- und Raumbereich darstellt, kann der Einfluss von Streuflächen auf das teilweise kohärente Licht bequem durch die Faltung der Wigner-Funktion mit der bi-directional scattering distribution function (BSDF) modelliert werden. Darüber hinaus zeigen wir, dass wir durch eine korrekte Definition der Lichtkohärenz die statistischen und deterministischen Oberflächenmodelle anwenden

können, um die Lichtstreuung von Hoch- und Mittelfrequenzfehlern gleichzeitig zu simulieren.

Abstract

The scattering of light deviates light from their desired trajectories in the optical system and can result in a degradation of the optical performances, mostly the contrast. Therefore, the simulation of light scattering in optical systems are vital for the system performance assessment. However, the current simulation methods of light scattering are mostly based on Monte Carlo raytracing, which is extremely time-consuming. Although there exist methods to accelerate the ray-based simulation of light scattering, their accuracy is limited, or they can hardly be applied for optical systems without circular symmetry. Furthermore, the ray-based methods are completely geometrical and can only simulate scattering due to surface roughness, while it cannot accurately model the small-angle scattering from mid-spatial frequency errors, which requires the modelling of the physical aspects of light such as partial coherence, diffraction and interference.

In this work, we first propose a quasi-analytical phase space method to simulate light scattering in optical systems, which directly couples the excitation and acceptance of light scattering in the phase space. Since the phase space method does not involve random sampling of the probability distribution function, it is much more efficient and delivers results with high signal to noise ratio (SNR). We first implement the phase space method to simulate the autofluorescence effect in microscope lenses, based on which a comprehensive analysis of the autofluorescence effect of various types of microscope lenses is presented. Subsequently, we extend the phase space method to simulate the surface scattering in optical systems. Meanwhile, we evaluate the extended phase space model in mirror systems with and without circular symmetry and propose three possibilities to implement the phase space method, which are suitable for different types of optical systems. By comparing the three implementations, the applicability, efficiency, accuracy and limitations of the phase space method are discussed.

Additionally, in order to simulate the propagation of partially coherent light in random media or by statistically perturbed surfaces, we propose a Wigner function-based method, in which the partial coherence of light is accurately modelled and the scattering of the partially coherent light by multiple scattering surfaces can be efficiently simulated. Since the Wigner function of light depicts the light distribution in the angular and spatial domain simultaneously, the impact of scattering surfaces on the partially coherent light can be conveniently modelled by the convolution of the Wigner function with the bi-directional scattering distribution function (BSDF) in the angular domain. Furthermore, we show that, by a proper definition of light coherence, we are able to apply the statistical and deterministic surface models to simulate light scattering from high and mid-spatial frequency errors simultaneously.

This page is intentionally left blank.

Contents

Zusammenfassung	i
Abstract	iii
Contents	v
1 Introduction and motivation	1
2 Fundamental Approaches to Scattering Simulation	4
2.1 Volume scattering.....	4
2.1.1 Rayleigh and Mie scattering.....	4
2.1.2 Autofluorescence of optical material.....	5
2.2 Surface scattering	6
2.2.1 Impact of surface errors on the image quality	6
2.2.2 K-correlation model.....	11
2.2.3 Harvey-Shack surface scatter theory	11
2.3 Monte Carlo raytracing.....	17
2.4 Importance sampling	19
2.5 The Wigner distribution	20
3 The phase space model of light scattering	23
3.1 Physical model.....	23
3.2 Quasi-analytical model.....	25
3.3 Geometrical surface scattering simulation	27
3.3.1 Surface discretization	28
3.3.2 Calculation methods of the acceptance angle.....	30
3.4 Geometrical volume scattering simulation.....	36
3.5 Field windows	37
3.6 Physical optics simulation based on the Wigner function.....	38
3.6.1 Ensemble average coherence.....	39
3.6.2 Propagation and scattering of partially coherent beams based on the Wigner function.....	42
4 Realization of Simulation Algorithms	48
4.1 Volume scattering.....	48
4.1.1 Description of the method	49
4.1.2 Simulation results and error analysis.....	51
4.2 Application and practical consequences in optical design	59
4.2.1 Overview	59

4.2.2	Autofluorescence coefficients of glass	61
4.2.3	Critical lens groups	64
4.2.4	Suppression of the autofluorescence effect in microscope lens design	69
5	Calculation of Surface Scattering.....	75
5.1	Geometrical phase space-based approach.....	75
5.1.1	Ritchey–Chrétien telescope.....	75
5.1.2	Kirkpatrick-Baez telescope	78
5.1.3	Critical review.....	83
5.2	Wigner function-based simulation of surface scattering.....	85
5.2.1	Retrofocus lens.....	85
5.2.2	EUV Schwarzschild objective	87
5.2.3	Limitations of the Wigner function method.....	90
6	Conclusions.....	93
	Appendix A: Isoenergetic sampling of incident rays by the Fibonacci grid	95
	References	97
	List of figures	100
	List of tables	106
	List of abbreviations.....	107
	List of symbols.....	109
	Acknowledgement.....	111
	Ehrenwörtliche Erklärung	112
	Publications	113

1 Introduction and motivation

The simulation of light propagation in optical systems is a vital step in optical design to assess the performance of an optical system before it is built. Typically, light is described by geometrical rays and the laws of refraction and reflection are used to model the propagation of the rays through optical systems. The geometrical ray tracing method is valid as long as the impact of the physical aspect of the light field, such as partial coherence, diffraction and interference, can be neglected. In contrary, in case the physical properties of the light field have a large impact on the performance of the optical system, the beam propagation methods that are based on the rigorous or approximated solutions of the Maxwell's equation should be used [1].

Besides the abovementioned geometrical and physical models of light propagation, another important phenomenon during light propagation is light scattering. There are various origins of light scattering in optical systems and their impact on the performance of the optical systems varies. The optical systems in which light scattering plays a dominating role are those that involve light propagation through turbid medium, such as biological tissues. In such systems, light is strongly scattered by the scatterers in the turbid medium and the modelling of diffusive light propagation through the turbid medium is necessary for the simulation of the whole optical system. On the other hand, for the optical systems that do not involve turbid medium, the scattering of light can result from the autofluorescence of the optical glass [2] and the surface scattering due to the residual surface roughness of the optical or mechanical surfaces [3].

Different from reflection and refraction, which can be analytically described by the laws of reflection and refraction, light scattering is a stochastic process due to the random distribution of scatterers or surface roughness. Therefore, light scattering results in random change of ray directions or random modulation of the wavefront, which complicates the simulation of light propagation. The commonly used methods to simulate light scattering are based on the Monte Carlo approach, in which the angular distributions of the scattered field are randomly sampled and consequently a large number of rays must be traced to accurately describe the scattered field.

In most of the commercial optical design softwares such as OpticStudio and FRED, light scattering is modelled by non-sequential Monte Carlo raytracing, which typically requires the tracing of millions or billions of rays to obtain an accurate result. Therefore, the straylight analysis based on the Monte Carlo simulation of light scattering is usually very time-consuming due to the large number of rays that need to be traced, and the signal to noise ratio (SNR) of the results is limited by the shot noise induced by random sampling. Furthermore, the Monte Carlo-based simulation of light scattering are completely ray-based and neglects several physical properties of light, in particular diffraction

effects. Therefore, the ray-based methods are not applicable when the scattering of partial coherent light or the diffraction of light from the mid-spatial-frequency (MSF) errors are of concern.

In order to tackle the abovementioned problems of the Monte Carlo-based methods in the simulation of light scattering, various methods have been developed to model different types of surface errors. For example, Harvey et. al. proposed an analytical method to estimate the impact of high frequency surface errors (micro roughness) on the transfer function of the optical system [4], Peterson proposed methods based on geometrical ray tracing [5], Liang et. al. proposed field propagation methods based on perturbation theory [6]. However, none of these methods is able to tackle statistical angular scattering and rigorous deterministic wave diffraction simultaneously nor deal with the propagation of partially coherent light in real optical systems with scattering surfaces. Furthermore, there lacks a unified model that is capable of modelling multiple scatterings from all frequency components of the PSD simultaneously.

Therefore, the main objective of this thesis is to develop an efficient simulation model for light scattering in optical systems, as well as a physical optics beam propagation method that models the multiple scattering of partially coherent light. In order to achieve this goal, a phase space model of light scattering has been proposed and implemented in the geometrical and physical methods, which greatly improves the efficiency of the simulation and enables the modelling of the physical properties of light in the simulation of light scattering.

In Chapter 2, a brief introduction on the models of light scattering is presented, including the Harvey-Shack surface scattering theory as well as the Rayleigh and Mie volume scattering models. Additionally, the modelling of partially coherent light by the Wigner function is introduced.

As the basis of this thesis, the phase space model is introduced in Chapter 3 together with its implementations in the simulation of volume and surface scattering, as well as the physical optics simulation of light scattering. For the simulation of volume scattering, the phase space model is implemented in the simulation of the autofluorescence effect of microscope lenses. By substituting volume discretization by surface discretization, the same method is then extended to the simulation of surface scattering in symmetry-free optical systems. Additionally, a quasi-analytical straylight model is presented, which provides insights into the geometrical generation and acceptance of the straylight in optical systems. In order to apply the Wigner function to simulate the scattering of partially coherent light, a combined definition of light coherence based on the ensemble and time average is introduced. Based on this combined definition, we are able to model the partial coherence, diffraction, interference, and multiple scattering of light simultaneously by the Wigner function method.

Examples of the applications of the phase space model in the simulation of volume scattering are presented in Chapters 4. As an example, the phase space method is applied to calculate the autofluorescence effect of a collection of microscope lenses, based on which a systematic analysis of the autofluorescence effect of microscope lenses is presented. The fundamental analysis reveals the key factors and critical lens groups for the autofluorescence effect. Based on these results, optical design and optimization strategies to suppress the autofluorescence effect in microscope lenses are discussed. Additionally, the accuracy of the phase space method in the autofluorescence simulation is validated against Monte Carlo raytracing and the limiting factors of the accuracy such as the paraxial approximations are discussed in the error analysis.

In addition to volume scattering, the geometrical and physical phase space-based surface scattering method is presented in Chapter 5. As examples, the geometrical approach is applied in the straylight analysis of a Ritchey-Chrétien telescope and a Kirkpatrick-Baez telescope. The Ritchey-Chrétien telescope is a circular symmetrical system in which it is shown that the impact of residual aberrations on the straylight distribution can be efficiently modelled by the Seidel coefficients. On the other hand, the Kirkpatrick-Baez telescope possesses extreme asymmetry and the method to calculate the acceptance solid angle greatly varies from that of circular symmetric systems. Depending on the symmetry of the optical system and the discretization of the detector, the phase space model can be implemented in three different methods for the geometrical simulation of surface scattering, and the efficiency, accuracy, applicability and robustness of the three methods are discussed in a critical review.

Additionally, as an example of the physical approach, the Wigner function method is applied in the propagation of partially coherent beams in an EUV Schwarzschild objective with two scattering mirrors. In the examples, the multiple scattering of light from the MSF and HSF errors are demonstrated and the PSF of the system is calculated.

2 Fundamental Approaches to Scattering Simulation

2.1 Volume scattering

Volume scattering occurs when the incident light impinges on a scatterer, which is a particle or an aggregate of particles whose optical properties are different from that of the surrounding medium. On a micro level, the probability of a photon being scattered by a single particle is measured by the particle's scattering cross-section δ_s . On a macro level, the scattering of light in a large bulk of medium is quantified by the scattering coefficient μ_s , which describes the probability of a photon being absorbed or scattered in unit propagation distance. Additionally, if the energy of the scattered photon is conserved after a scattering event, it is called *elastic scattering*, while *inelastic scattering* corresponds to a scattering event in which the energy of the scattered photon is not conserved. In this section, we briefly discuss the origins and modelling of these two types of volume scattering events in optical systems.

2.1.1 Rayleigh and Mie scattering

The Rayleigh and Mie scattering models have been widely used to model elastic scattering by scatterers with different sizes.

Rayleigh scattering happens when the size of the scatterer is much smaller than the wavelength of the incident photon (up to 1/10 of the wavelength). It results from the electric polarizability of the particles. The oscillating electric field of the light wave creates an oscillating dipole, when the electrons are moved back and forth. The oscillating dipole itself then becomes a radiator whose radiation is the scattered light. According to the Rayleigh scattering model, the scattering phase function of a spherical scatterer with refractive index n and diameter d for an incident wavelength of λ is [7]

$$P(\theta) = \frac{1}{8} \frac{\pi^4 d^6}{\lambda^4} \left| \frac{n^2 - 1}{n^2 + 2} \right|^2 (1 + \cos^2 \theta), \quad (2.1)$$

From the above equation, we can see that Rayleigh scattering has a strong dependence on the size of the scatterer and the wavelength of the incident light, while the scattered light intensity is symmetrical in the forward and backward directions.

When the size of the scatterer grows larger and become comparable to the wavelength of the incident light, Mie scattering becomes dominant. The term Mie scattering, also known as the Mie solution, describes the scattering of an electromagnetic plane wave by stratified spheres or cylinders. The derivation of Mie solution starts from the Maxwell's equation and the boundary conditions on the surface of the sphere. In spherical polar coordinates, the equations separate into a set of ordinary differential equations, which are then solved

for the two subfields in the form of infinite series [8]. Fig. 2-1 shows the phase function calculated by the Mie solution, from which we can see that the symmetry of the scattered intensity in the forward and backward direction is lost when the size of the scatterer is comparable to the wavelength and forward scattering is dominant.

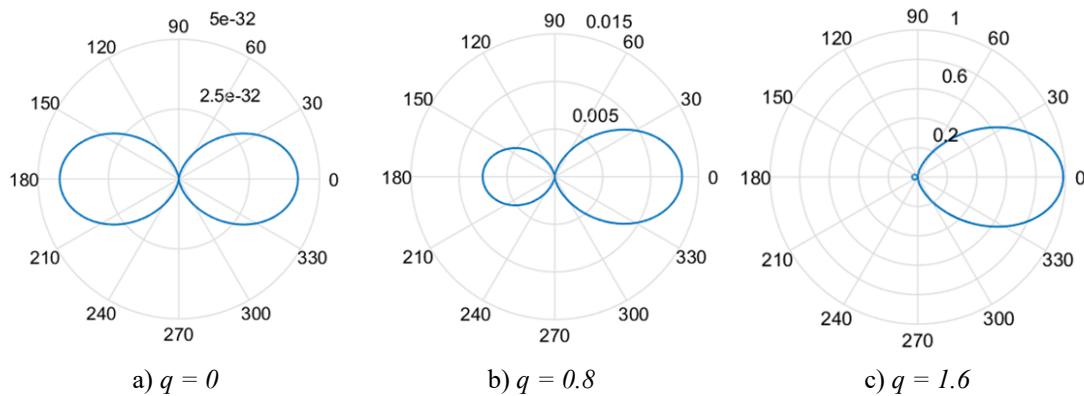


Fig. 2-1. Polar diagrams of the scattering phase function of a spherical scatterer with a refractive index of $n = 1.33$ calculated by the Mie solution, $q = 2\pi/\lambda$.

2.1.2 Autofluorescence of optical material

The autofluorescence of optical material is an inelastic scattering process which occurs when the optical material is illuminated by excitation light. Typically, autofluorescence occurs in plastic optical materials such as PMMA or in optical glass when it is illuminated with ultraviolet light.

The autofluorescence of optical glass is particularly critical in fluorescence microscopy, in which UV light is commonly used as excitation light [9]. Due to the relatively weak fluorescence signal emitted by the fluorophores in the biological sample, the straylight that originate from the autofluorescence of optical glasses has a large impact on the contrast of the image and reduces the SNR. An important type of scatterers for the autofluorescence of optical glasses is the heavy metal atoms or lanthanide series of chemical elements. Fig. 2-2 shows relative autofluorescence coefficients of some optical glasses for an excitation wavelength of 365nm (i-line), from which we can see that the lanthanum glasses as well as the lead-containing glasses (non-N type) show the highest autofluorescence coefficient [10].

The autofluorescence light spectrally overlaps with the sample related signal and cannot be filtered easily. In an ideal objective for fluorescence microscopy, materials should have very low autofluorescence levels. The choice of a low-fluorescing microscope objective is important for high-sensitivity imaging because sufficiently low fluorescent background has to be obtained for the detection of single molecules. Autofluorescence emitted by an objective has to be minimized in the design phase of an optical instrument through judicious material selection.

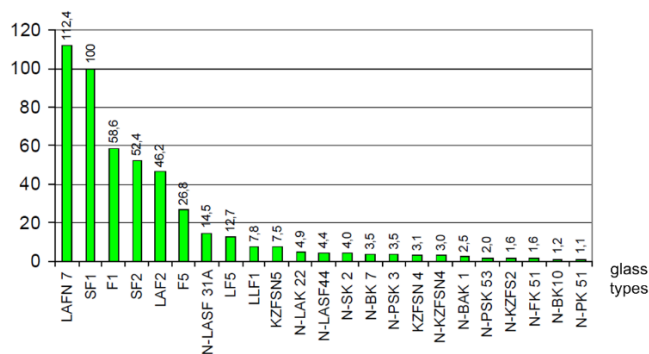


Fig. 2-2. Relative fluorescence coefficients of some Schott glasses normalized on the value of SF1. The excitation is at $\lambda = 365$ nm and the signal is spectrally integrated [10].

2.2 Surface scattering

2.2.1 Impact of surface errors on the image quality

Light is scattered at optical surfaces due to the existence of manufacturing errors or surface contamination. In optical design, the scattering due to the residual manufacturing error are the main origin of straylight and therefore must be taken into account in surface tolerancing and straylight analysis. The surface height variations induced by manufacturing errors are highly irregular and modelled as stochastic processes. In particular, the surface errors are usually represented in the Fourier domain by the power spectral density (PSD), which is the Fourier transform of the autocovariance function of the surface profile. The PSD provides a quantitative evaluation of the frequency components contained in the residual surface errors. As shown in Fig. 2-3, the residual surface errors can be categorized into three types, which are figure error, mid-spatial frequency (MSF) error and high frequency surface roughness. Examples of the three types of surface errors are shown in Fig. 2-4. The figure errors are of low spatial frequency and correspond to the deviation of the surface from the desired shape. The surface roughness is the high frequency components of the PSD which originate from the polishing process. Between the figure error and micro roughness, the MSF errors often contain intrinsic regular structures that originate from the manufacturing process, such as the concentric grooves on diamond-turned surfaces or vibrations of the fabrication machines. The two peaks in the PSD shown in Fig. 2-3 corresponds to the regular structures induced by surface grinding.

Each of the frequency component of the PSD has different impact on the performance of optical surfaces. The figure errors are usually represented by the Zernike polynomials and induces additional aberrations to the optical system. The surface roughness scatterers light to a large angular range, resulting in a non-zero offset on the image plane that decreases the contrast. The scattering behavior of MSF errors are more complicated, since the

correlation length of the MSF errors is usually comparable to the correlation length of light, and thus the diffraction effect of the MSF errors should be considered. Furthermore, regular periodic ripples act as a phase grating and creates structured false light. Typically, the MSF errors results in small angle scattering that broadens the point spread function (PSF) and decreases the resolution of the system. More specifically, in the following paragraphs, we will discuss the impact of different types of surface errors on the PSF and modulation transfer function (MTF) of optical systems.

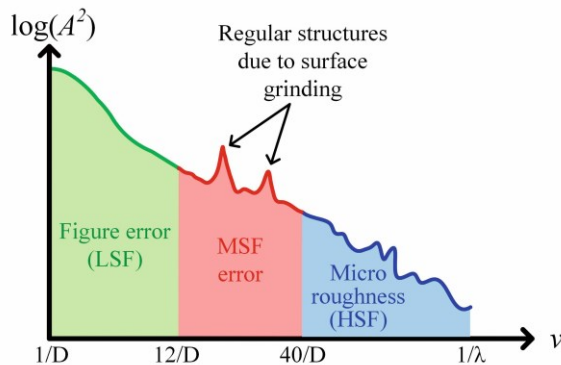


Fig. 2-3. Log-log scaled power spectral density (PSD) of optical surfaces manufactured by surface grinding. The figure error is the low spatial frequency (LSF) part while the micro roughness is the high spatial frequency (HSF) part of the PSD.

As we know, monochromatic light emitted by a point object is fully coherent. Therefore, if the surface errors do not vary with time, the light field at the exit pupil is also fully coherent, even though it contains extremely complex phase variations due to the residual surface errors. Consequently, the amplitude distribution at the image plane can be considered as the coherent superposition of all the Huygens wavelets from the exit pupil.

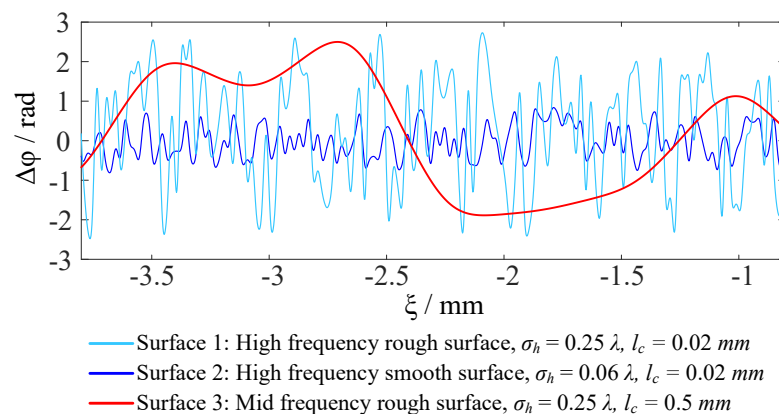


Fig. 2-4. Illustration of the surface height variation of three types of surface errors. The surface variations are modelled as Gaussian processes with different correlation lengths and RMS surface height. Only part of the full aperture is shown for clear illustration.

As shown in Fig. 2-5, the complex field at the exit pupil is given by the intrinsic pupil function $P_i(\zeta, \eta)$ and the space-variant random phase deviation $P_s(\zeta, \eta)$ induced by the residual errors. The intrinsic part of the pupil function is inherent to the optical design and can be obtained by raytracing or more rigorous beam propagation methods, while the

impact of the residual surface error on the complex pupil function is more complicated due to the presence of mid and high spatial frequency surface errors.

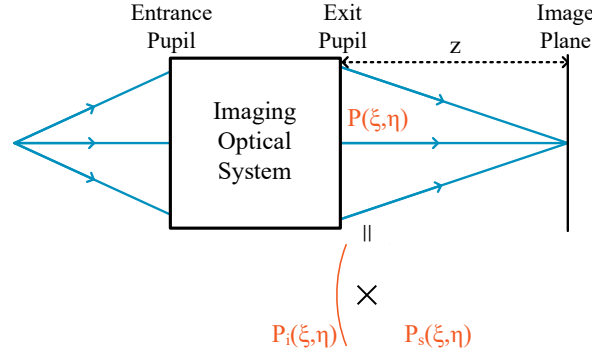


Fig. 2-5. Schematic sketch of an imaging optical system with scattering surfaces, $P_i(\xi, \eta)$ is the intrinsic pupil function that includes all the geometrical aberrations, apodizations and edge diffractions, while $P_s(\xi, \eta)$ is the additional pupil function that encapsulates the phase variation at the exit pupil induced by the residual surface errors.

According to the scalar diffraction theory, the amplitude at the image plane can be written as the Fourier transform of the total complex pupil function:

$$A(x, y; z) = \frac{e^{j\phi(x, y, z)}}{\lambda z} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P_i(\xi, \eta) P_s(\xi, \eta) \exp \left[-j \frac{2\pi}{\lambda z} (x\xi + y\eta) \right] d\xi d\eta \quad (2.2)$$

Applying the convolution theorem of Fourier transform, the amplitude at the image plane can then be written as the convolution between the amplitude spread function produced by the intrinsic pupil function and the residual surface errors.

$$A(x, y; z) = A_i(x, y; z) \otimes A_s(x, y; z) \quad (2.3)$$

If the residual surface error is modelled as a random process, its impact on the complex pupil function is that it adds random phase shifts to each of the Huygens wavelets that constitute the amplitude distribution function $A_s(x, y; z)$ on the image plane. Therefore, $A_s(x, y; z)$ is proportional to the Fourier transform of a series of random phasors, which produces speckle-like distributions and broadens the energy distribution on the image plane. To illustrate the impact of such broadening effect, we add the surface errors shown in Fig. 2-4 separately to the pupil function of the system shown in Fig. 2-5. The three surface error profiles are modelled by zero-mean random process with different standard deviation $\hat{\sigma}_h$ and correlation length l_c . The optical system is assumed to be well corrected with zero aberration and apodization. With these surface profiles the exact PSFs are calculated by taking the product of (2.2) with its complex conjugate and the results are shown in Fig. 2-6. Fig. 2-6(a) shows the irradiance distribution near the central peak, which is significantly attenuated by the surface errors. Fig. 2-6(b) shows the irradiance distribution in a larger range of spatial coordinates, from which we can observe that the spatial spreading of the PSFs is increased by the surface errors. Combining with Fig. 2-6(c), which shows the irradiance distribution at even larger spatial coordinates, we can observe that the high spatial frequency errors tend to scatter the light to larger angles,

resulting in reduction of image contrast, while MSF errors mainly contribute to small angle scattering, leading to reduction of the resolution. The impact of the standard deviation and the correlation length of surface error on the spatial distribution of scattered light can also be seen from encircled energy shown in Fig. 2-6(d), from which we can observe that the MSF errors mostly scatter light to small scattering angles.

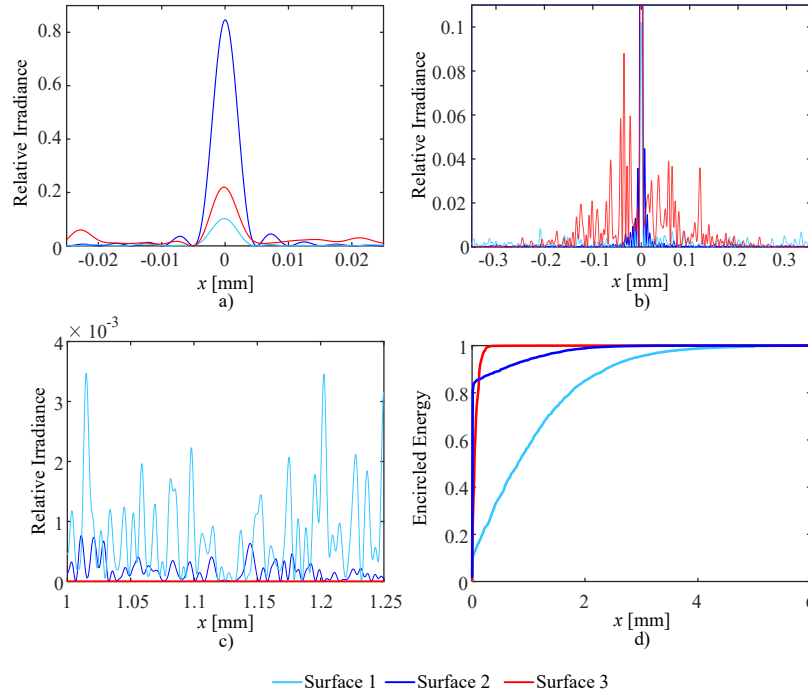


Fig. 2-6. 1-dimensional PSFs of optical systems that contains the surface errors of Fig. 2-4. a) shows the PSFs near the central peak. b) shows the PSFs in a larger spatial range. c) shows the side lobes of the PSFs far from the central peak. d) shows the encircled energy of the PSFs.

While the results shown in Fig. 2-6 provides insights into the scattering properties of different types of surface errors, it should be noted that surface 1 and 3 are optically rough surfaces with their standard deviations comparable to one wavelength, which is realistic for MSF errors, but not common for high frequency surface roughness. In reality, optical surfaces should be polished to be optically smooth, with their standard deviations of surface roughness much smaller than a wavelength ($\sigma_h \ll \lambda$). As it can be observed from the PSF of the smooth surface 2 in Fig. 2-6(a), the sinc^2 distribution of the PSF is preserved near the central peak, while the main impact of the small surface roughness (micro roughness) is that it scatters a small fraction of energy (15%) to larger spatial range. Therefore, for optically smooth surfaces, calculation of the exact intensity distribution is of little practical use. Instead, the degradation of image contrast, which is directly related to the transfer function, should be analyzed with statistical methods. Harvey et. al have already proven that if the surface error can be modelled by a zero-mean Gaussian process with wide-sense stationarity, the impact of the surface error on the transfer function $H_s(u, v)$ can be decoupled with the intrinsic transfer function $H_i(u, v)$, and the mathematical expectation of the overall transfer function $H(u, v)$ of the optical system is written as [4]:

$$H(u, v) = H_s(u, v) \cdot H_i(u, v) \quad (2.4)$$

where $H(u, v) = E \left\{ \exp \left[j2\pi (h_1 - h_2) \right] \right\}$, $E \{ \}$ denotes mathematical expectation and \hat{h}_1, \hat{h}_2 are the surface heights at two random locations normalized by the wavelength.

It should be noted that the transfer function defined by (2.4) is not the exact transfer function of a particular optical system but the ensemble average over different systems with the same design and statistical properties of optical surfaces, the difference between the exact and average transfer function can be seen in Fig. 2-7, from which we can observe that the exact MTF of the smooth surface 2 oscillates closely around the predicted MTF, with only minor deviations.

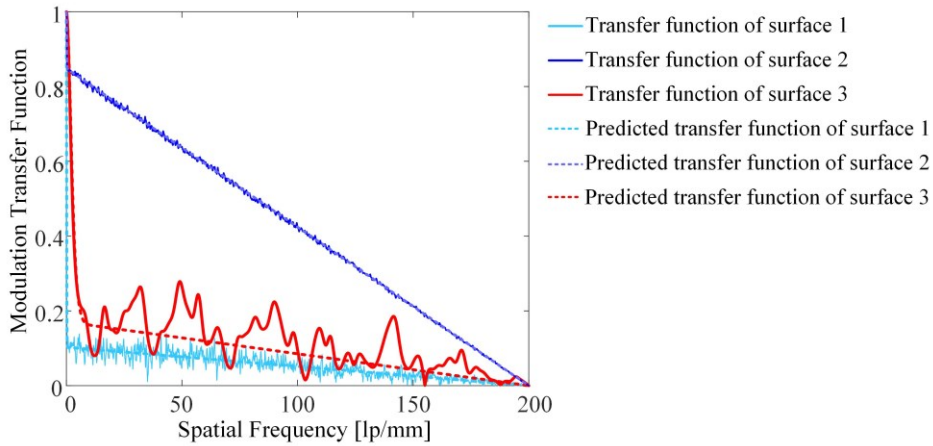


Fig. 2-7. Modulation transfer function (MTF) of the optical systems with residual surface errors. The solid lines correspond to the actual MTFs of the surface error 1-3 shown in Fig. 2-4. The dashed lines indicate to the MTFs predicted by the Harvey model.

Therefore, the exact MTF induced by micro roughness is of little use and the average MTF predicted by the Harvey model is adequate to characterize the image quality degradation, which is a constant ratio of contrast decrease for all spatial frequencies. However, the same conclusion cannot be drawn for MSF errors such as surface 3, whose exact MTF oscillates slowly around the predicted MTF but with deviations that are so large that the exact MTF falls to 0 at spatial frequencies smaller than the cut-off frequency, which corresponds to a reduction of resolution. This corresponds to our previous discussion about the small angle scattering effect of mid-spatial frequency errors. Furthermore, it should be noted that the profile of surface 3 is modelled by a random process with large correlation lengths, while for real optical surfaces the MSF errors cannot be completely modelled by random spatial variations due to the presence of the regular structures resulting from the manufacturing processes. Such regular structures of surface error prohibit the application of statistical methods on the modelling of such errors. Therefore, in order to predict the image quality degradation due to MSF errors, the exact surface metrology data of MSF structure is required, based on which the exact MTF or PSF should be calculated.

2.2.2 K-correlation model

As discussed in the last section, the residual surface errors of optical surfaces are represented by the PSD. In order to model the PSD in the light scattering simulations, an analytical form of the PSD is desired. If we assume a Gaussian autocovariance function given by

$$ACV = \sigma_s^2 \exp\left[-(r/l_c)^2\right], \quad (2.5)$$

the PSD, which is the Fourier transform of the ACV , is also Gaussian and given by

$$PSD(f_x, f_y) = \pi \sigma_s^2 l_c^2 \exp\left[-\pi^2 l_c^2 (f_x^2 + f_y^2)\right] \quad (2.6)$$

Gaussian PSDs are mathematically convenient to be implemented in surface scattering models, but in reality, most optical surfaces do not have a Gaussian autocovariance function and PSD. Instead, typical PSDs tend to have an inverse power law falloff at larger spatial frequencies [11]. A more general model that better describes this behavior is the ABC or K-correlation model [12], in which the two-dimensional PSD is given by

$$PSD(f_x, f_y) = \frac{1}{2\sqrt{\pi}} \frac{\Gamma(s/2)}{\Gamma[(s-1)/2]} \frac{AB}{\left[1 + B^2 (f_x^2 + f_y^2)\right]^{s/2}}, \quad (2.7)$$

where s is the slope of the 2D PSD, $B = 2\pi l_c$ when $s = 3$, and Γ is the well-known Gaussian Gamma function. The normalization factor A ensures that the integration of the PSD gives the effective surface roughness variance

$$\sigma_{eff}^2(\lambda) = \int_{1/L}^{1/\lambda} PSD(f_x, f_y) df_x df_y. \quad (2.8)$$

2.2.3 Harvey-Shack surface scatter theory

In order to predict the scattering property based on its PSD, or to retrieve the PSD of a surface from the measured scattering distribution, a connection between the PSD and the angularly resolved scattering distribution must be established. For this purpose, the Harvey-Shack (HS) surface scattering theory is most commonly used [4].

The HS theory starts with the Rayleigh-Sommerfeld diffraction integral. As shown in Fig. 2-8, consider an incident monochromatic collimated beam along the direction (α_i, β_i) , where α_i and β_i are the direction cosines of the incident beam in the x and y directions. The incident field is diffracted by a non-absorbing reflective diffracting aperture with a diameter of d and phase modulation of $p_R(x', y')$. The diffracted wave field on a hemisphere with a radius of curvature of r centered at the center of the diffracting aperture is given by

$$U(\alpha, \beta; \hat{r}) = \gamma \frac{e^{i2\pi\hat{r}}}{\hat{r}} \iint_{\Sigma'} U_o(x', y') \left[\frac{1}{2\pi\hat{r}(1+\varepsilon)} - i \right] \frac{1}{(1+\varepsilon)^2} e^{i2\pi(\hat{l}-\hat{r})} dx' dy', \quad (2.9)$$

where $U_o(x', y')$ is the field distribution right after the diffracting aperture, $\alpha = x/\hat{r}$, $\beta = y/\hat{r}$, $\gamma = \hat{z}/\hat{r}$ are the direction cosines of the observation point, $\varepsilon = (\hat{l} - \hat{r})/\hat{r}$, and the caret symbols correspond to length quantities scaled by the wavelength.

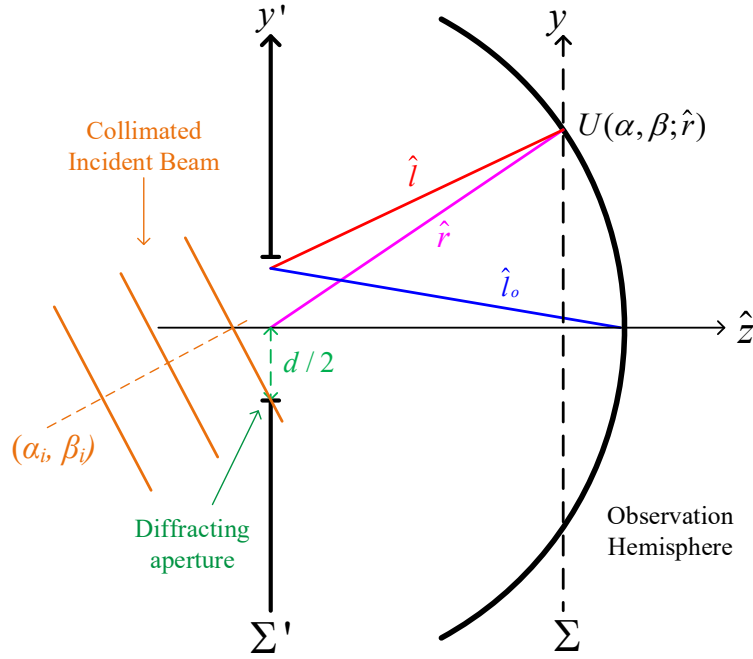


Fig. 2-8. Observation of the diffracted field by a diffracting aperture [13].

Assuming that the radius of curvature of the observation hemisphere is much larger than the wavelength and the diameter of the diffracting aperture, we have $\hat{r} \gg 1$ and $d \ll \hat{r}$. In this case, applying the Fraunhofer approximation and we can simplify Eq. (2.9) as

$$U(\alpha, \beta; \hat{r}) = \gamma \frac{e^{i2\pi\hat{r}}}{i\hat{r}} \iint_{\Sigma'} U_o(x', y') e^{-i2\pi(\alpha x' + \beta y')} dx' dy'. \quad (2.10)$$

The above equation indicates that the field distribution on the observation hemisphere is merely a Fourier transform of the field at the diffracting aperture

$$U(\alpha, \beta; \hat{r}) = \gamma \frac{e^{i2\pi\hat{r}}}{i\hat{r}} \text{FT} [U_o(x', y')]. \quad (2.11)$$

Therefore, the scattered flux within the solid angle $d\Omega$ subtended by $d\alpha$ and $d\beta$ is

$$d\phi = \gamma \text{FT} [U_o(x', y')]^2 d\Omega, \quad (2.12)$$

and the irradiance intensity in the direction (α, β) is given by

$$I(\alpha, \beta) = \gamma FT \left[U_o(x', y') \right]^2. \quad (2.13)$$

In order to characterize the dependence of the irradiance distribution $I(\alpha, \beta)$ on the incident wave and the phase modulation of the diffracting aperture, we can rewrite the field right after the diffracting aperture as follows

$$U_o(x', y') = a_0 p_L(x', y') p_R(x', y'), \quad (2.14)$$

where a_0 is the amplitude of the incident light, $p_L(x', y')$ is the phase distribution of the incident beam and $p_R(x', y')$ is the phase modulation of the diffracting aperture, which are given by the following equations

$$\begin{aligned} p_L(x', y') &= \exp \left[i2\pi \left(x'\alpha_i + y'\beta_i \right) \right] \\ p_R(x', y') &= \exp \left[i4\pi W_R(x', y') \right], \end{aligned} \quad (2.15)$$

where $W_R(x', y')$ is the optical path variation induced by the thin scattering element at the diffracting aperture in waves.

Now let us consider the bidirectional reflectance distribution function (BSDF) of the diffracting aperture. The BSDF is a 4D function of the incident and reflection direction cosines. For a particular incident direction, the BSDF degenerates into a 2D function of the reflection direction cosines, and it is defined as the radiance of the reflected light along the diffraction cosine (α, β) over the incident irradiance E_i . For the case shown in Fig. 2-8, the BSDF can be written as

$$\begin{aligned} \text{BSDF}(\alpha_i, \beta_i, \alpha, \beta) &= \frac{L(\alpha, \beta)}{E_i} \\ &= \frac{I(\alpha, \beta)}{\gamma A_s \cdot E_i} \\ &= \frac{I(\alpha, \beta)}{\gamma A_s \cdot a_0^2} \\ &= \frac{1}{A_s \cdot a_0^2} \cdot FT \left[U_o(x', y') \right]^2. \end{aligned} \quad (2.16)$$

Applying the convolution theorem of Fourier transform and substituting Eq. (2.14) into Eq. (2.16), we get

$$\text{BRDF}(\alpha_i, \beta_i, \alpha, \beta) = \frac{1}{A_s} FT \left\{ \left[p_L(x', y') p_R(x', y') \right] \otimes \left[p_L(x', y') p_R(x', y') \right]^* \right\}, \quad (2.17)$$

where \otimes denotes convolution. From Eq. (2.17) we can see that the BSDF is given by the Fourier transform of the autocorrelation of the pupil function. Considering Eq. (2.15), we can write the autocorrelation of the pupil function as follows

$$\begin{aligned} & \left[p_L(x', y') p_R(x', y') \right] \otimes \left[p_L(x', y') p_R(x', y') \right]^* \\ &= \iint_{A_s} \exp \left[i2\pi (u\alpha_i + \hat{v}\beta_i) \right] \exp \left\{ i4\pi \left[W_R(x', y') - W_R(x'-u, y'-\hat{v}) \right] \right\} dx' dy', \end{aligned} \quad (2.18)$$

where u and \hat{v} are the pupil displacement in waves. In order to investigate the scattering property of a scattering surface, we assume that the diffracting aperture shown in Fig. 2-8 is a part of a scattering surface. In this case, considering the fact that the surface height variations of a scattering surface is highly irregular, an analytical form of $W_R(x', y')$ does not exist and therefore the integration in Eq. (2.18) cannot be analytically calculated. However, if we define the BSDF as the average over an ensemble of scattering surfaces on which the surface height variations share the same statistical properties, we are able to calculate the analytical form of the BSDF, by which the statistical dependence of the surface scattering properties on the surface height variations is revealed. Taking the ensemble average of Eqs. (2.17) - (2.18), we get the ensemble average definition of the BSDF

$$\begin{aligned} & E \langle BRDF(\alpha_i, \beta_i, \alpha, \beta) \rangle \\ &= \frac{1}{A_s} FT \left\{ E \left\langle \left[p_L(x', y') p_R(x', y') \right] \otimes \left[p_L(x', y') p_R(x', y') \right]^* \right\rangle \right\} \\ &= \frac{1}{A_s} FT \left\{ \exp \left[i2\pi (u\alpha_i + \hat{v}\beta_i) \right] \iint_{A_s} E \left\langle \exp \left\{ i4\pi \left[W_R(x', y') - W_R(x'-u, y'-\hat{v}) \right] \right\} \right\rangle dx' dy' \right\}. \end{aligned} \quad (2.19)$$

where $E \langle \rangle$ denotes ensemble average. If $W_R(x', y')$ corresponds to the surface height variations resulting from the microroughness of optical surfaces, the following conditions are usually fulfilled:

- i. $W_R(x', y')$ is a single-valued Gaussian random process.
- ii. $W_R(x', y')$ is locally stationary (surface is homogeneous and isotropic).
- iii. The correlation length of $W_R(x', y')$ is small enough such that the random variables at two fixed pair of spatial coordinate $W_R(x', y')$ and $W_R(x'-u, y'-\hat{v})$ are jointly normal.
- iv. $W_R(x', y')$ is weakly ergodic.

In this case, the mathematical expectation $E \left\langle \exp \left\{ i4\pi \left[W_R(x', y') - W_R(x'-u, y'-\hat{v}) \right] \right\} \right\rangle$ can be calculated by the joint characteristic function $\Phi(\omega_1, \omega_2)$, where ω_1 and ω_2 are the

constant factor that scales the two random variables [14]. Therefore, the mathematical expectation can be calculated as follows:

$$\begin{aligned} E \left\langle \exp \left\{ i4\pi \left[W_R(x', y') - W_R(x' - u, y' - \hat{v}) \right] \right\} \right\rangle \\ = \Phi(4\pi, -4\pi) \\ = \exp \left\{ - \left(4\pi\sigma_w \right)^2 \cdot \left[1 - C_w(u, \hat{v}) / \sigma_w^2 \right] \right\}, \end{aligned} \quad (2.20)$$

where σ_w^2 and $C_w(u, \hat{v})$ are the deviation and the autocovariance of the surface height variation $W_R(x', y')$.

For well-polished optical surfaces, the surface height variations are usually much smaller than a wavelength and we have $\sigma_w^2 \ll 1$ and $C_w(u, \hat{v}) \ll 1$. For such smooth surfaces, we can apply the smooth surface approximation to further simplify Eq. (2.20)

$$E \left\langle \exp \left\{ i4\pi \left[W_R(x', y') - W_R(x' - u, y' - \hat{v}) \right] \right\} \right\rangle = 1 - 16\pi^2\sigma_w^2 + 16\pi^2C_w(u, \hat{v}). \quad (2.21)$$

Substituting Eq. (2.21) into Eq. (2.19), and recall that the Fourier transform of the autocovariance gives the PSD, we get the ensemble average of the BSDF

$$\begin{aligned} E \langle \text{BSDF}(\alpha_i, \beta_i, \alpha, \beta) \rangle \\ = (1 - \text{TIS}) \delta(\alpha - \alpha_i, \beta - \beta_i) + \frac{16\pi^2}{\lambda^2} \text{PSD} \left(\frac{\alpha - \alpha_i}{\lambda}, \frac{\beta - \beta_i}{\lambda} \right), \end{aligned} \quad (2.22)$$

where $\text{TIS} = 16\pi^2\sigma_w^2$ is the total integrated scattering, which corresponds to the fraction of the total reflected radiant power that is scattered out of the specularly reflected beam. Therefore, we can see from Eq. (2.22) that the radiance of the reflected light is composed of two part, one being the specularly reflected part along the same direction as the specular beam, the other being the diffusively reflected part which depend linearly on the PSD of the scattering surface. Additionally, the diffusive part of the BSDF is also known as the angular spread function (ASF)

$$\text{ASF}_{\text{OHS}}(\alpha_i, \beta_i, \alpha, \beta) = \frac{16\pi^2}{\lambda^2} \text{PSD} \left(\frac{\alpha - \alpha_i}{\lambda}, \frac{\beta - \beta_i}{\lambda} \right). \quad (2.23)$$

Eq. (2.23) is also known as the original Harvey-Shack (OHS) surface scatter model [4]. According to this model, the ASF is shift-invariant in the directional cosine space for different incident angles. The OHS theory can predict the angular distribution of the scattered light in small scattering angles. However, due to the neglect of the incident and scattering angle in Eq. (2.15), the OHS theory is not accurate for large incident and scattering angles. In order to increase the accuracy of the OHS for large incident angles,

the phase modulation of the diffracting aperture $p_R(x', y')$ in Eq. (2.15) is modified to project the surface height to the direction of the incident light

$$p_R(x', y') = \exp\left[i4\pi \cos \theta_i W_R(x', y')\right], \quad (2.24)$$

where θ_i is the incident angle of the incident beam. Substituting Eq. (2.24) into Eqs. (2.19) - (2.23), we get the ASF with an additional dependence on the incident angle

$$\text{ASF}_{\text{MHS}}(\alpha_i, \beta_i, \alpha, \beta) = \frac{16\pi^2}{\lambda^2} (\cos \theta_i)^2 \text{PSD}\left(\frac{\alpha - \alpha_i}{\lambda}, \frac{\beta - \beta_i}{\lambda}\right). \quad (2.25)$$

This is the so-called modified Harvey-Shack (MHS) surface scattering model. When the incident angle is very small, the MHS model degenerates into the OHS model. Due to the projection of the surface height in Eq. (2.24), the MHS model is able to accurately predict the ASF for large incident angles. However, due to the additional dependence on the incident angle, the MHS model is no longer shift-invariant.

Although the MHS tackles the problem of large incident angles, it fails to accurately predict the ASF for large scattering angles due to the fact that the phase modulation induced by the surface height variations is also dependent on the scattering angle. As shown in Fig. 2-9, for an incident angle of θ_i and a scattering angle of θ_s , the phase modulation due to the surface height $h(x', y')$ can be written as

$$W_R(x', y'; \theta_i, \theta_s) = 2\pi(\cos \theta_i + \cos \theta_s)h(x', y') \quad (2.26)$$

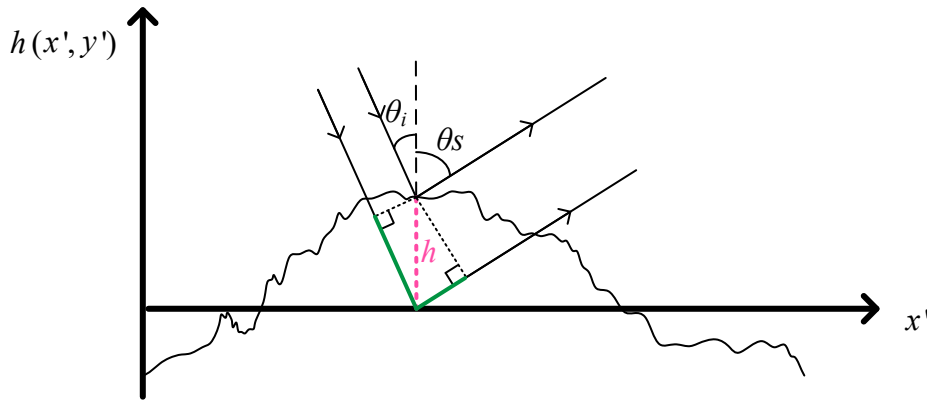


Fig. 2-9. Phase modulation by surface height variations of a scattering surface for oblique incidence and scattering angle. The green segments correspond to the optical path difference (OPD) induced by the surface height h .

Substituting Eq. (2.26) into Eqs. (2.19) - (2.23), we get the ASF for finite incident and scattering angles

$$\text{ASF}_{\text{GHS}}(\alpha_i, \beta_i, \alpha, \beta) = \frac{16\pi^2}{\lambda^2} (\cos \theta_i + \cos \theta_s)^2 \text{PSD}\left(\frac{\alpha - \alpha_i}{\lambda}, \frac{\beta - \beta_i}{\lambda}\right). \quad (2.27)$$

The above equation is also known as the generalized Harvey-Shack (GHS) surface scattering model. Taking the incident and scattering angle into account, the GHS model is able to predict the ASF for large incident and scattering angles.

The Harvey-Shack surface scattering models reveal the connection between the PSD of optical surfaces with its BSDF. Combining the K-correlation model for the surface PSD discussed in Sec. 2.2.2, we can write the ASF of polished optical surfaces as follows

$$\begin{aligned} \text{ASF}_{\text{GHS}}(\alpha_i, \beta_i, \alpha, \beta) \\ = \frac{16\pi^2}{\lambda^2} (\cos \theta_i + \cos \theta_s)^2 \frac{1}{2\sqrt{\pi}} \frac{\Gamma(s/2)}{\Gamma[(s-1)/2]} \frac{AB}{[1 + B^2(f_x^2 + f_y^2)]^{s/2}}, \end{aligned} \quad (2.28)$$

where

$$\begin{aligned} \cos \theta_i &= \sqrt{1 - \alpha_i^2 - \beta_i^2} \\ \cos \theta_s &= \sqrt{1 - \alpha_s^2 - \beta_s^2} \\ f_x &= \frac{\alpha - \alpha_i}{\lambda} \\ f_y &= \frac{\beta - \beta_i}{\lambda}. \end{aligned} \quad (2.29)$$

Eqs. (2.28) - (2.29) provides an analytical form of the ASF for real optical surfaces and they have been widely used in the inverse scattering problem to characterize optical surfaces by measured ASFs. In this work, we use Eq. (2.28) to model the ASFs or BSDFs of optical surfaces.

2.3 Monte Carlo raytracing

Classically, the simulation of light scattering in optical systems is based on Monte Carlo raytracing. In the Monte Carlo method, a large number of rays are launched into the optical system from the predefined light sources. The initial locations and directions of the rays are determined by the properties of the light source, and random sampling in the angular and spatial domain is usually needed to model the size and NA of the light source. Fig. 2-10 shows the simulation of volume and surface scattering by non-sequential Monte Carlo raytracing in OpticStudio.

After the rays are launched into the optical system, they undergo refraction and reflection at the optical surfaces, and most importantly, when the rays reach the scattering surface or volume, it has to be determined whether or not the rays are scattered. The chances that the rays are scattered in the optical system is determined by the scattering properties of the scattering surface or volume. For example, for scattering surfaces, the probability that a ray is scattered is determined by the TIS, while for a scattering volume, the probability

is governed by the scattering coefficient and the path length that the ray travels within the scattering volume. If a ray is scattered, the direction of the scattered ray has to be determined. For scattering volumes, the scattering direction can be determined by random sampling of the scattering phase function, which are discussed in Sec. 2.1. For surface scattering the new direction of the ray is determined by randomly sampling the BRDF of the surface, as discussed in Sec. 2.2.

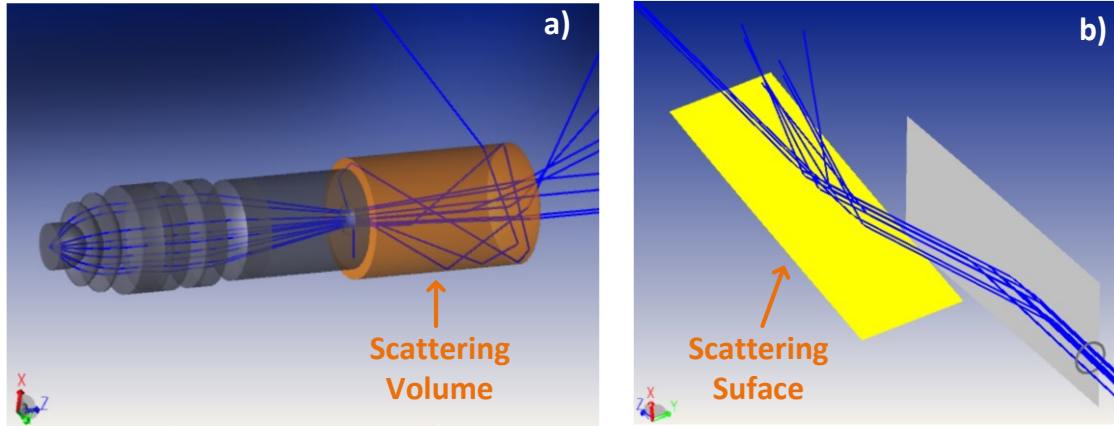


Fig. 2-10. Simulation of volume scattering in a microscope lens (a) and surface scattering in a Kirkpatrick-Baez telescope (b) by Monte Carlo raytracing in OpticStudio. As an illustration, 10 rays are launched into the optical systems from the on-axis object in each system.

The most efficient way to determine the scattering angle is to randomly sample the cumulative distribution function. For example, if the probability density distribution function of the scattering angle is $P(\theta)$, the scattering angle θ_s can be determined by

$$\xi = \int_0^{\theta_s} P(\theta) d\theta, \quad (2.30)$$

where ξ is a random number uniformly distributed between $[0,1]$. The advantage of this sampling method is that it statistically biases the scattering direction according to the scattering angular distribution, so that each ray carries the same flux.

In order to apply Eq. (2.30) for the sampling of the scattering angle, the analytical form of the integral of the probability density distribution function is required. However, the integrals of the scattering phase functions and ASFs predicted by Mie and GHS theories cannot be expressed analytically. Consequently, the volume scattering phase functions are usually approximated by the Henyey-Greenstein (HG) model, whose integral can be given analytically. For surface scattering, the typical approach to determine the scattering angle is to scatter rays uniformly to the hemisphere, and then assign different weighting to the scattered rays according to the ASF. However, due to the uniform sampling of the rays on the hemisphere, a large fraction of rays is scattered to large scattering angles, which correspond to small weightings considering the very small values of the ASF for large scattering angles. Because tracing these rays with small weightings consumes significant computation power while makes limited contribution to the straylight

information, modelling the ASF by ray weightings further reduces the efficiency of the Monte Carlo method. Another alternative to sample the non-analytically-integrable scattering distribution functions is to use iterative random sampling, which requires many iterations to generate a scattered ray if the scattering is highly directional. Therefore, the modelling of the angular distribution of the scattered rays is a limiting factor for the accuracy and efficiency of the Monte Carlo method.

In the final step of the Monte Carlo simulation, the flux of the rays that hit the detector are summed up to calculate the straylight distribution on the detector. Since the flux carried by each ray is superposed incoherently, there is a presumption that the scattered field is completely incoherent. This presumption is true for the scattering from microroughness due to the very small coherence length of such surface errors, while it is not true for surface errors with larger correlation lengths such as the MSF errors. Therefore, for the simulation of surface scattering, the Monte Carlo method is in principal only valid for the scattering from microroughness. While for the surface scattering from MSF errors, the interference of the scattered field from different parts of the surface should be considered due to the finite degree of spatial coherence. The spatial coherence of scattered light will be further discussed in Sec. 3.6.

2.4 Importance sampling

Since the application of the Monte Carlo method in light scattering simulation is strongly limited by its efficiency, various methods have been proposed to accelerate the simulations. Among these methods, importance sampling is the most efficient one which has minor impact on the accuracy of the simulation.

As discussed in the last section, the low efficiency of the Monte Carlo method mainly originates from the large number of rays that are traced. Fig. 2-11 shows the Monte Carlo-based straylight analysis for a three-mirror anastigmat (TMA) system with and without important sampling. From Fig. 2-11(a) we can observe that if the FOV of the optical system is small and large angle scattering occurs, most of the scattered rays are irrelevant to the straylight distribution on the detector because they are blocked or propagated out of the optical systems before reaching the detector. Consequently, tracing these irrelevant rays yields no information of the straylight while consumes significant computational power.

As shown in Fig. 2-11(b), importance sampling directs most of the scattered rays to the detector, so that the tracing of the irrelevant rays can be avoided. In this case, we are able to obtain results with high signal-to-noise ratio (SNR) by tracing a small number of rays, and the efficiency of the Monte Carlo method can be significantly increased. The acceleration ratio of importance sampling is strongly dependent on the distribution of the

ASF and the FOV of the system. Typically, high acceleration ratio is expected for optical system with wider ASF and smaller FOV.

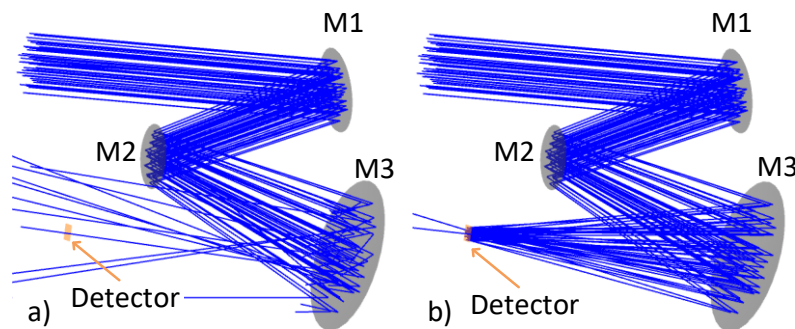


Fig. 2-11. Monte Carlo raytracing without importance sampling (a) and with importance sampling (b) for a TMA system. The target object of importance sampling is the detector. M3 is a scattering surface with a Gaussian BRDF.

The disadvantage of important sampling is that it requires additional effort to predetermine the area of interest, to which the scattered light should be directed. In the case shown in Fig. 2-11, the area of interest is the detector because M3 is the last surface in the system. However, calculation of the straylight contributions from M1 and M2 are not very straightforward, because there are other elements between these two surfaces and the detector. Therefore, the area of interest for M1 and M2 should be determined by finding the intermediate image of the detector seen by these two surfaces. As a consequence, the user must locate the intermediate images of the detector by an initial raytracing before setting up importance sampling for the Monte Carlo simulation. In some optical design softwares, the determination of the area of interest can be automatized by user-defined scripts, but these methods are mostly restricted to circularly symmetric systems, in which paraxial raytracing and the Lagrange invariant can be applied to estimate the location and size of the detector intermediate image [15]. While in non-circularly symmetric systems, importance sampling is usually set up manually. Furthermore, since only part of the complete scattered ray cone is modelled by important sampling, the raytracing result on the detector must be scaled radiometrically by the BRDF to obtain the actual straylight distribution.

2.5 The Wigner distribution

As discussed in Sec. 2.3, the partial coherence of the scattered light should be considered if the scattering surface has a large correlation length. Therefore, it is necessary to model the partial coherence of light during the light scattering simulation.

The common practice to model partially coherent light is through modal expansion [16]. In this method, a partially coherent beam is expanded to a series of mutually incoherent eigenmodes, which are propagated individually. The representation of partially coherent light by coherent modes enables us to propagate partially coherent beams with physical

optics beam propagators, but it is rather computationally intensive considering the large amount of eigenmodes required to represent the scattered beam with low degree of coherence. Furthermore, if there are multiple scattering surfaces in the optical system, the coherent eigenmodes become partially coherent after each scattering event, therefore, the modal expansion has to be repeated after every scattering surface, which further increases the computational complexity if multiple scattering is involved.

A more efficient method to model partially coherent light is the Wigner function. For a partially coherent light field $E(x)$, its autocorrelation function is defined as

$$\Gamma(x, \Delta x) = \left\langle E\left(x + \frac{\Delta x}{2}\right) E^*\left(x - \frac{\Delta x}{2}\right) \right\rangle, \quad (2.31)$$

where angle brackets denote time average.

The Wigner function of the partially coherent light field is then defined as the Fourier transform of its autocorrelation function [17]

$$W(x, u) = \int \Gamma(x, \Delta x) \exp(-i \frac{2\pi}{\lambda} u \Delta x) d\Delta x. \quad (2.32)$$

From the above equation we can see that the Wigner function depicts the distribution of light in the spatial and angular domain simultaneously, which is especially advantageous for the modelling of light scattering because the scattering of light is usually described by an angular extension of the incident light.

Another advantage of the Wigner function is that it can be propagated conveniently by the ABCD matrix under the paraxial approximation. According to the theory of linear optical transformations, the direction and position of an incident ray (x, u) is related to that of the output ray (x', u') by the matrix transformation:

$$\begin{pmatrix} x' \\ u' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix}. \quad (2.33)$$

Therefore, we have

$$x = \frac{Bu' - Dx'}{BC - AD}; u = \frac{Cx' - Au'}{BC - AD}, \quad (2.34)$$

and the transformation of the Wigner function by the ABCD matrix is given by

$$W'(x', u') = W\left(\frac{Bu' - Dx'}{BC - AD}, \frac{Cx' - Au'}{BC - AD}\right), \quad (2.35)$$

where the prime symbol denotes the light field leaving the optical system represented by the ABCD matrix. As an example, Fig. 2-12 shows the propagation of a partially coherent top-hat Gaussian-Schell beam through a positive lens followed by a free-space propagation. From Fig. 2-12(b, c) we observe that the perfect lens and free-space

propagation result in shearing of the Wigner function in the u and x directions correspondingly.

Furthermore, the propagation of the Wigner function through thin elements is also straightforward. Assume that the phase and amplitude modulation of the thin element can be described by $t(x)$, we can write the light field after a thin element as

$$E'(x) = t(x)E(x). \quad (2.36)$$

Substituting Eq. (2.36) into Eqs. (2.31) and (2.32), and applying the convolution theorem of Fourier transform, we get the Wigner function after the thin element

$$\begin{aligned} W'(x', u') &= \int \left\langle E\left(x + \frac{\Delta x}{2}\right) E^*\left(x - \frac{\Delta x}{2}\right) t\left(x + \frac{\Delta x}{2}\right) t^*\left(x - \frac{\Delta x}{2}\right) \right\rangle \exp(-i \frac{2\pi}{\lambda} u \Delta x) d\Delta x \quad (2.37) \\ &= \int \Gamma(x, \Delta x) \Gamma_t(x, \Delta x) \exp(-i \frac{2\pi}{\lambda} u \Delta x) d\Delta x, \end{aligned}$$

where Γ_t is the cross-spectral density function of the thin element. Therefore, the propagation of a partially coherent beam through a thin element can be conveniently modelled by the Wigner function.

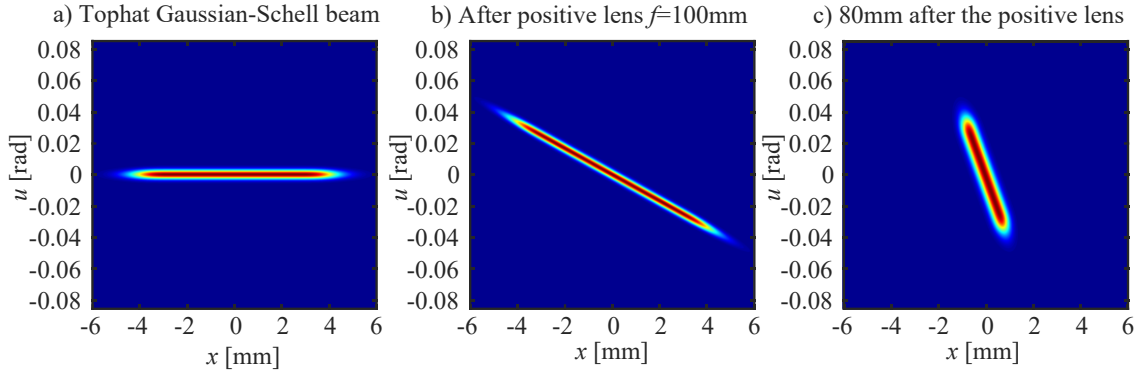


Fig. 2-12. Wigner functions of a top-hat Gaussian-Schell beam after a positive lens and free-space propagation.

3 The phase space model of light scattering

In the last chapter, the limitations of the existing methods for light scattering simulations have been discussed, and the aim of this work is to develop new physical models and simulation methods to overcome these limitations in the efficiency, accuracy and applicability. To achieve this goal, we propose a method in which the scattering of light is modelled in the phase space. The phase space model depends on the etendue-matching raytracing from the forward and backward direction to determine the generation and acceptance of the straylight, which are coupled quasi-analytically in the phase space. The etendue-matching raytracing significantly saves computational cost because it allows us to skip the rays that have no chance to reach the detector. Therefore, the phase space method is advantageous in both efficiency and accuracy. Furthermore, the phase space coupling of the straylight provides more insights into the physical nature of light scattering and straylight generation. In this chapter, we first introduce this phase-space-based physical model, and subsequently discuss the implementation of the model in geometrical and physical optics simulation methods.

3.1 Physical model

In general, the straylight in optical systems due to light scattering is generated in two steps, which are the scattering of the incident light and acceptance of the scattered light. In the first step, part of the scattering volume or surface is illuminated by the incident light, which is then scattered by the scatterers or surface roughness within the area of illumination. The light scattering effect can be interpreted as a redistribution of the incident light in the angular domain. In the acceptance step, the scattered light is propagated through the rest of the optical system to determine the fraction of the scattered light that is accepted by the optical system. Therefore, the key to determine the straylight distribution is to calculate the overlapping of the scattered light and the acceptance capability of the optical system in the spatial and angular domain.

Since the phase space simultaneously describes the distribution of light in the spatial and angular domain, it can conveniently describe the two steps of straylight generation mentioned above. As an illustration, Fig. 3-1 shows a simple example of 2-dimensional straylight generation. As shown in Fig. 3-1(a), a quasi-collimated incident beam with a diameter of d is scattered by an infinitely large scattering surface, which scatters the incident light to a larger angular range as indicated by the green arrows. The phase space distributions of the incident and scattered light are depicted in Fig. 3-1(b,c). The scattered light leaving the scattering surface is then truncated by an aperture stop with a diameter of a , and part of the truncated scattered light is propagated to the detector AB . In order to determine the acceptance of the detector in phase space, two pairs of rays are traced from the points A and B through the edges of the aperture stop to the scattering surface. As

indicated in Fig. 3-1(a,d), the phase space acceptance of the detector for straylight is determined by the angles and positions of the reversely traced rays. Finally, the total flux of scattered light on the detector can be calculated as

$$\Phi_{AB} = \int_{-d/2}^{d/2} \int_{-\pi/2}^{\pi/2} S(y,u)A(y,u)dydu, \quad (3.1)$$

where $S(y,u)$ and $A(y,u)$ are the phase space distributions of the scattered light and detector acceptance. The above integral can be conveniently calculated by the overlapping between the scattered light and detector acceptance in the phase space, as indicated by dark red area in Fig. 3-1(e).

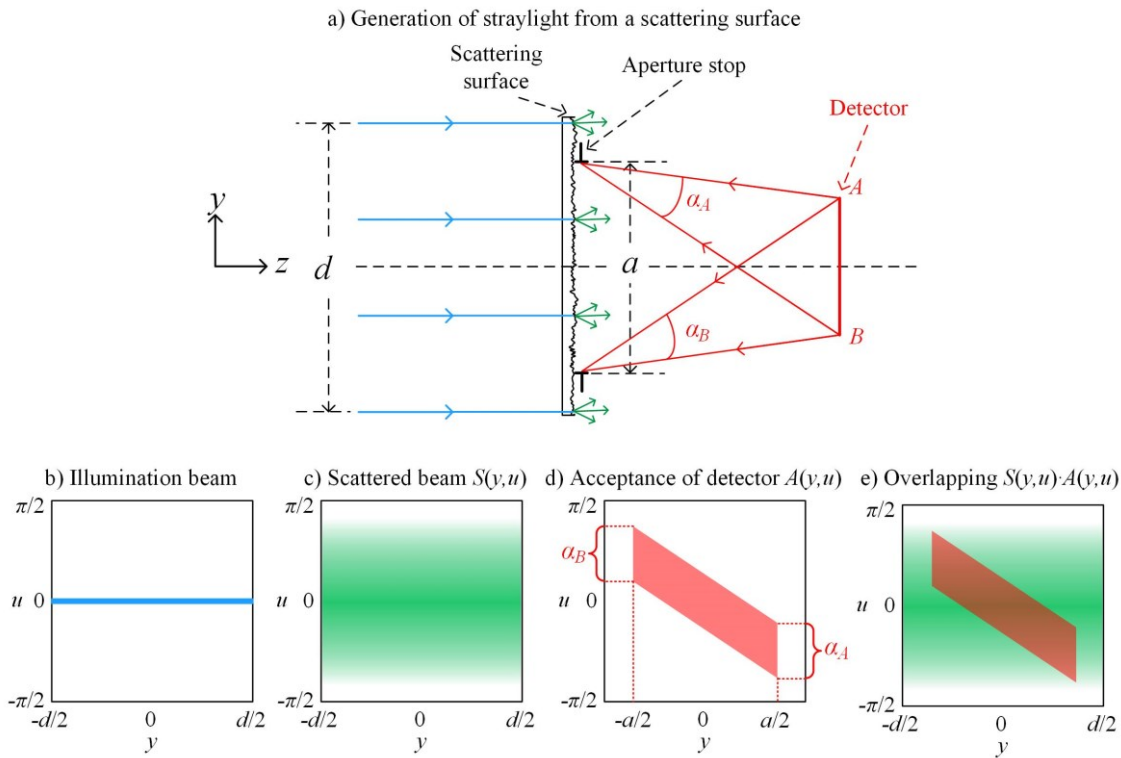


Fig. 3-1. Illustration of the generation of straylight from a scattering surface. a) shows the layout of the setup. b) and c) show the phase space distribution of the illumination and scattered light. d) shows the acceptance of the detector in the phase space. e) shows the overlapping between the generation and acceptance of scattered light.

Due to the convenient coupling of the generation and acceptance of scattered light in the phase space, the phase space model is very efficient in the calculation of straylight. More specifically, as shown in Fig. 3-1(a), we need to trace two marginal rays from the object space to determine the area of illumination, and trace two pairs of coma rays from the upper and lower edges of the detector to determine the area and angle of acceptance. Therefore, in total we only need to trace 6 rays to calculate the straylight flux on the detector, which results in great advantage in efficiency compared to the Monte Carlo raytracing. Furthermore, since no random sampling is involved, the phase space model delivers results with high SNR. In particular, if the straylight distribution on the detector

rather than the total flux is required, the Monte Carlo method typically suffers from the shot noise due to the fine sampling of the detector, while for the phase space model, one simply has to repeat the integral in Eq. (3.1) for each pixel on the detector to obtain the straylight distribution with high SNR.

3.2 Quasi-analytical model

Although the physical picture of the phase space model can be clearly illustrated by the example shown in Fig. 3-1, it is far from being practical because scattering surfaces in real optical systems are usually separated from the detector by additional optical components, and additionally the integral in Eq. (3.1) should be calculated in 3D. In this section, we extend our discussion to real optical systems and establish a quasi-analytical model which reveals the key parameters of an optical element that determines its straylight contribution.

From Fig. 3-1(d), we observe that the shape of the phase space acceptance is a parallelogram, which means that the acceptance angle is constant in the acceptance area, this is true if the scattering surface is close to the aperture stop and the paraxial approximation is valid. Additionally, if we assume that the detector or pixel is very small such that the angular distribution of the scattered light is constant within the acceptance angle, we can further simplify Eq. (3.1) as

$$\Phi_{AB} = \alpha \cdot \int_{-d/2}^{d/2} S(y)A(y)dy, \quad (3.2)$$

where α is the acceptance angle of the detector. Additionally, if we extend the calculation to 3D, we have

$$\Phi_{AB} = d\Omega \cdot \iint_{\Sigma} S(x, y)A(x, y)dxdy, \quad (3.3)$$

where Σ is the illuminated area and $d\Omega$ is the solid acceptance angle which is related to the acceptance angle α as

$$d\Omega = 2\pi(1 - \cos\alpha). \quad (3.4)$$

In aperture-dominant systems such as microscope objectives or telescopes with small FOV, the integral in Eq. (3.3) is nearly constant through the optical system due to the large overlapping between the illumination and acceptance area. Therefore, for such systems the acceptance solid angle $d\Omega$ is the determining factor for the straylight contribution from an element, and an analytical model to describe the dependency of the acceptance solid angle on various parameters is necessary.

Here we restrict our discussion to aperture-dominant systems and consider a simplified model of a microscope lens shown in Fig. 3-2. The simplified model is composed of a front group and a middle group that decreases the NA, and a rear group that collimates

the ray bundle and a tube lens that forms an image. The detector has a half-diameter of p_i , and plane O is an arbitrary pupil plane inside the lens. We assume the chief ray height at this plane to be 0, which is justified by the fact that this is an aperture-dominant system.

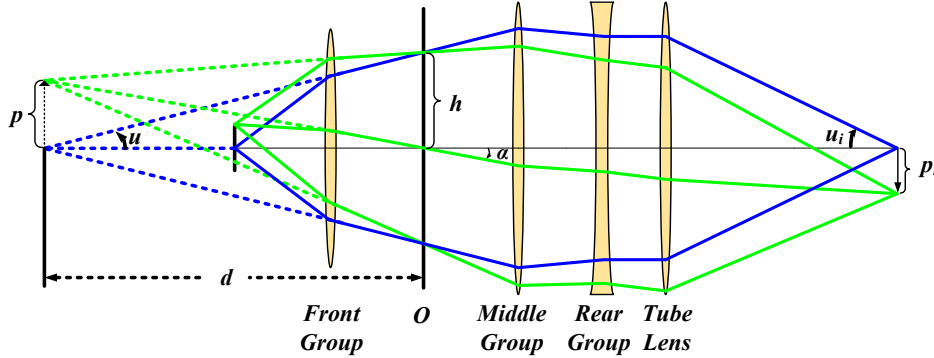


Fig. 3-2. Sketch of a simplified model of a microscope lens.

The acceptance solid angle $d\Omega_o$ of plane O is characterized as the angular opening of the detector as seen from plane O, which equals the angular opening of the intermediate image of the detector as seen from plane O. The radius of the intermediate image is denoted by p and the 2D angular opening is α . Additionally, we have $\tan \alpha = p/d = pu/h$. Considering the Lagrange invariant $npu = n_i p_i u_i$, we have $\tan \alpha = n_i p_i u_i / nh$. By trigonometric transformation we get

$$d\Omega_o = 2\pi \left(1 - \frac{1}{\sqrt{\tan^2 \alpha + 1}} \right). \quad (3.5)$$

If α is small enough, $d\Omega_o = \pi\alpha^2$, $\alpha = p/d$, we get

$$d\Omega_o = \pi p_i^2 (n_i u_i)^2 \frac{1}{n^2 h^2}. \quad (3.6)$$

Therefore, the acceptance angle is determined by the refractive index of the element, the marginal ray height and $\pi p_i^2 (n_i u_i)^2$, which is the etendue of the system, while it is completely independent on the aperture angle inside the element. This is similar to the analytical surface scattering model derived by Peterson [5].

Though the above derivation seems to be simple and explicit, note that it is only valid in the paraxial range. For example, in typical microscope lenses, the aperture angle is small in the middle and rear groups, which means that the paraxial model is, in general, applicable in these two lens groups. However, the actual value of the acceptance angle might differ from the analytical solution due to the finite aperture angles.

An example of the violation of the paraxial approximation is the front group of the microscope lenses, in which the aperture angle can be large due to the large NA. Fortunately, in most of the microscope lenses the front groups are composed of quasi-aplanatic meniscus lenses, in which the sine condition is closely fulfilled. Therefore, it gives us another opportunity to overcome the paraxial limit and analytically calculate the real acceptance angle in case the sine condition is fulfilled.

Assuming that the sine condition is fulfilled, we can substitute the Lagrange invariant in the previous derivation by the sine condition

$$np \sin u = n_i p_i \sin u_i \quad (3.7)$$

Assuming the acceptance angle α to be very small, we get

$$\Delta\Omega_{o\text{-sine}} = \pi p_i^2 (n_i \sin u_i)^2 \frac{1}{n^2 h^2} \cdot \frac{1}{\cos^2 u} \quad (3.8)$$

Therefore, in case the aperture angle is large and the sine condition is fulfilled, the acceptance solid angle is also dependent on the aperture angle inside the element, larger aperture angles yield larger acceptance solid angles.

3.3 Geometrical surface scattering simulation

In the last sections, we have simplified the phase space coupling in the 2D phase space, which is more efficient for optical systems with circular symmetry. It is easier to illustrate but is based on certain approximations. On the other hand, for real optical system, especially for systems without circular symmetry, the straylight contribution from an element should be evaluated in the 4D phase space with the dimensions x , y , u and v . Furthermore, the illumination intensity and acceptance solid angles are usually not uniformly distributed on an optical surface. Therefore, the overlapping of the 4D phase space distribution should be evaluated in the spatial and angular domain separately. Consequently, discretization of the scattering surface is necessary and the angular straylight coupling should be evaluated within each subarea of the scattering surface.

Fig. 3-3 shows the determination of the acceptance angle in a surface subarea. In this figure, the red ray indicates the illumination rays from the object, which determines the amount of energy received by the subarea dS of the optical surface, while the ratio of the scattered energy to the received energy is determined by the BRDF of the surface together with the solid angle $d\Omega$ within which the scattered rays can reach the detector.

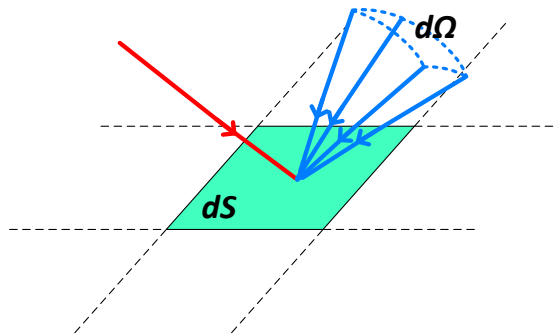


Fig. 3-3. Coupling of the illumination and acceptance in a subarea of a reflective surface. The red ray is traced from the object space while the blue rays are traced from the four corners of the detector. The pentahedron cone formed by the four blue rays defines the solid acceptance angle.

Based on the BRDF of the surface and the solid acceptance angle of acceptance $d\Omega$, the ratio of the scattered energy that reaches the detector over the incident energy in a surface subarea can be calculated by integrating the BRDF over the solid acceptance angle $d\Omega$:

$$\phi_s / \phi_i = \iint_{d\Omega} BRDF(\theta_i, \varphi_i, \theta_s, \varphi_s) \cos \theta_s d\theta_s d\varphi_s \quad (3.9)$$

where ϕ_i is the flux of the incident light and $\theta_i, \varphi_i, \theta_s, \varphi_s$ are the deflection and azimuthal angles of the incident and scattered rays.

According to the definition of the BRDF in Eq. (2.16), when the incident angle is fixed, the BRDF becomes a 2D function of θ_s and φ_s . The angles of the rays from the four corners of the detector then form a quadrangle in the $\theta_s - \varphi_s$ space, within which the integration of Eq. (3.9) can be calculated. Furthermore, for optical systems with a small field of view (FOV), or when a single pixel on the detector is considered, the solid angle $d\Omega$ is usually so small that the value of BRDF is considered to be constant within the acceptance angle. In this case, the integration of Eq. (3.9) can be further simplified as:

$$\phi_s / \phi_i = BRDF(\theta_i, \varphi_i, \theta_s, \varphi_s) \cdot \cos \theta_i \cdot d\Omega. \quad (3.10)$$

Based on Eq. (3.10), the irradiance on a pixel of the detector contributed by one surface subarea can be calculated as:

$$\begin{aligned} E_{pixel} &= \frac{S_{sub}}{S_{pixel}} \cdot E_{sub} \cdot BRDF(\theta_i, \varphi_i, \theta_s, \varphi_s) \cdot \cos \theta_i \cdot d\Omega \\ &= \frac{S_{sub}}{S_{pixel}} \cdot E_{sub} \cdot \eta_c, \end{aligned} \quad (3.11)$$

where E_{sub} is the irradiance of the incident light in the surface subarea, S_{sub} is the area of the subarea of the optical surface, S_{pixel} is the area of the pixel in the detector intermediate image and $\eta_c = BRDF(\theta_i, \varphi_i, \theta_s, \varphi_s) \cdot \cos \theta_i \cdot d\Omega$ is the coupling coefficient that corresponds to the fraction of the incident flux scattered to the pixel.

3.3.1 Surface discretization

In order to obtain a meaningful trade-off between accuracy and computational complexity, it is more advantageous to divide the optical surface into equal-area subareas (isoenergetic sampling). The type of grid used for surface discretization depends on the geometry of the surface. For example, a Cartesian grid is always isoenergetic, but it is not able to accurately describe curved boundaries, and therefore is usually used to describe non-circularly symmetric surfaces. On the other hand, a special isoenergetic polar grid is used to describe the circularly symmetric surfaces with the smallest number of sampling points. Fig. 3-4 shows an optical surface with a Gaussian irradiance distribution which is indicated by the colour code. As can be seen from the red mesh, an

isoenergetic polar grid with only 7 rings (49 subareas) is able to describe this distribution with good accuracy, and the small number of subareas used for surface discretization greatly reduces the number of rays that need to be traced in the phase space simulation.

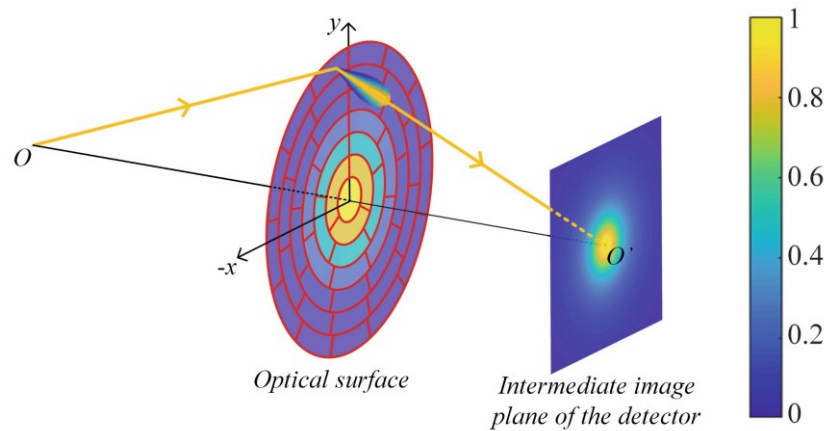


Fig. 3-4. Generation of straylight in a subarea of an optical surface. O is an on-axis object point and O' is on the intermediate image plane of the detector (details to follow in Sec. 5.1). The red mesh on the optical surface indicates the isoenergetic polar grid used to discretize the surface, and the color code indicates the relative irradiance distribution on the optical surface due to Gaussian apodization. The color code on the intermediate image plane of the detector indicates the contribution of the surface subarea to the irradiance distribution on the detector.

Similar to the sampling of the optical surfaces, the sampling of the entrance pupil should also be isoenergetic. In this case, the Cartesian grid is again not the best option because it induces artefacts due to the Moiré effect, especially when Cartesian discretization of the surface is applied. Therefore, it is better to sample the entrance pupil on a Fibonacci grid, which avoids the artefacts while guaranteeing uniform sampling. The details of the Fibonacci sampling grid are discussed in Appendix A.

For both Cartesian and isoenergetic sampling of the surface, the sampling density is dependent on the uniformity of the distribution of irradiance and solid acceptance angle. If both quantities are uniformly distributed on the surface, only a few subareas are enough to characterize the scattering property of the whole surface. It should be noticed that the number of necessary subareas to reach a requested accuracy grows with the intensity gradients on the surface. The uniformity of the irradiance distribution depends on the apodization, vignetting and aperture angle of the ray bundle inside the optical system. Additionally, the projection effect of large field angles also gives rise to non-uniformity of the irradiance distribution. In general, for aperture-dominant systems with a small FOV and a large aperture (e.g., typical telescope systems), the irradiance distribution is rather uniform throughout the system and dense sampling of the target surface is usually not required. A quantitative evaluation of the impact of sample density on the accuracy is given in Sec. 5.1.3.

The discretization of optical surfaces also allows us to apply space-variant BSDFs to different surface locations, which is useful when the inhomogeneity of residual surface

errors is of concern. As shown in Fig. 3-4, the straylight contribution from a surface subarea is calculated based on the irradiance and BSDF at this surface location.

3.3.2 Calculation methods of the acceptance angle

In optical systems with large FOV, the acceptance angle of the detector is spatially-variant. Therefore, the acceptance angle should be determined in each subarea individually. The most straightforward way to determine the acceptance angle is to trace rays reversely from the corners of the detector to the subareas, which is very efficient for paraxial systems that can be described by the ABCD matrix. However, for optical systems with large aberrations or complicated geometry, tracing rays reversely to a specific subarea involves real ray aiming, whose accuracy and efficiency are limited by the residual aberration of the optical elements between the subarea and the detector. Alternatively, the acceptance solid angle for a subarea can also be determined by the position and size of the intermediate image of the detector as seen by the subarea. In this case, the key to calculate the acceptance solid angle is the determination of this detector intermediate image. Under the paraxial approximation, the position of the detector intermediate image can be easily obtained by finding the intersection point of the paraxial marginal ray with the optical axis, and the size of the detector can be calculated with the Lagrange invariant. However, for general optical systems with large apertures or non-circularly symmetric systems, the presence of residual aberrations and loss of circular symmetry distorts the detector intermediate image and complicates the calculation of the acceptance solid angle. More importantly, since each subarea on the optical surface corresponds to different marginal ray heights, each subarea sees a different detector intermediate image due to the pupil-dependence of aberrations, as shown in Fig. 3-5.

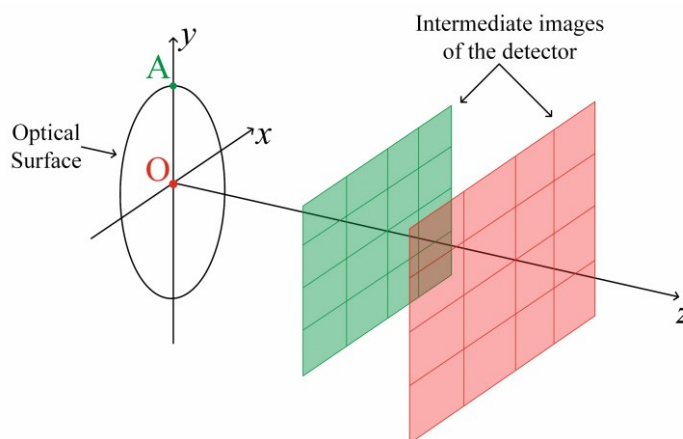


Fig. 3-5. Intermediate images of the detector seen by different subareas in the presence of spherical aberrations. The green mesh is the intermediate image seen by the off-axis point A, while the red curve is that seen by the on-axis point O.

Considering the complexity of the calculation of the acceptance solid angle, three different methods to calculate the acceptance solid angle for different types of optical systems have been proposed. In this section, we introduce the physical principals of three

methods, while a detailed comparison of the methods will be given in Sec. 5.1.3 after the presentation of two examples.

3.3.2.1 Intermediate image method

The first method is based on the aberrated intermediate image of the detector seen by the optical surface. Real imaging optical systems are usually imperfect due to the presence of aberrations, which result in a distorted, blurred and curved intermediate image of the detector. Moreover, because different subareas on an optical surface correspond to different pupil coordinates, and most aberrations except for distortion vary with the pupil coordinates, each subarea on the optical surface sees a unique aberrated intermediate image determined by the corresponding pupil coordinates of the subarea. Therefore, in order to determine the exact position of a pixel in the aberrated detector intermediate image, we need to consider the pupil coordinates defined by the distance of the subarea to the chief ray, as well as the field coordinate defined by the position of the pixel in the detector. Fig. 3-6 illustrates the relationship between the pupil and field coordinates with the corresponding subarea and pixel locations. Once the pupil and field coordinates are determined, the pixel position shift can be calculated from the transverse aberration. There are multiple ways to determine the amount of aberrations that the intermediate image suffers from, but the most efficient method is to calculate in 3rd order approximation the Seidel sums of the part of optical system between the target surface and the image plane. Based on the Seidel sums, the transverse aberration of a pixel can be calculated as follows:

$$\Delta x = -\frac{R}{n} \frac{\partial W}{\partial x_p}; \Delta y = -\frac{R}{n} \frac{\partial W}{\partial y_p}, \quad (3.12)$$

where W is the Seidel representation of the wavefront error, y_p is the absolute pupil coordinate, R is the radius of the reference sphere and n is the refractive index in the intermediate image space.

There are several advantages in using the Seidel sums, the first being that they can be easily calculated from the surface parameters or by tracing a few primary rays, which makes this approach extremely efficient. The second advantage is that the field dependence of the aberrations is explicitly given, therefore we only need to calculate the Seidel coefficients for one field and the aberrations for other field points can be directly obtained. The disadvantage of the Seidel representation is that it does not consider higher order aberrations, which leads to errors when those aberrations are present. In principle, for well corrected circularly symmetric systems, it is usually sufficient to consider only the primary aberrations since the coefficients of high order aberrations are typically orders of magnitudes lower. But in a sophisticated system involving freeform surfaces, it might be necessary to consider the impact of higher order aberrations on the detector

intermediate image. In this case, we can use the Zernike coefficients to represent the aberrations in the intermediate image space, but the calculation of the Zernike coefficients requires denser pupil sampling and the tracing of more rays, therefore being less efficient. Additionally, the scalability of the Zernike coefficients with field should be handled with care because different orders of the Zernike coefficients can have different field-dependence.

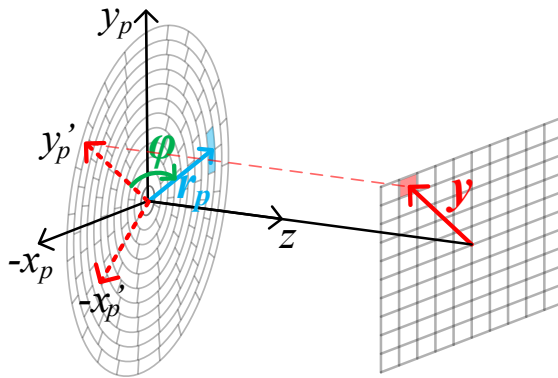


Fig. 3-6. The field coordinate y and pupil coordinates (r_p, ϕ) of an arbitrary subarea-pixel combination. The red and blue areas indicate the pixel and surface subarea. For simplicity, the chief ray is assumed to intersect with the optical surface at the optical axis, while in general cases, the pupil coordinates should be referred to the actual intersection point of the chief ray with the surface. For each pixel, the transverse aberration is evaluated in the local coordinate system in which the y axis is parallel to the field vector as indicated by the red dashed line.

The impact of aberrations on the detector intermediate image largely depends on the type of aberration. As mentioned above, the simplest case is distortion because it is independent of the pupil coordinates, meaning that if only distortion is present, every subarea of the optical surface sees the same distorted intermediate image of the detector. In this case, the positions of the off-axis pixels are shifted radially, inducing a radial variance of the pixel size that must be considered when the solid angles subtended by the pixels are calculated. Fig. 3-7(f) shows the distorted detector grid in the presence of distortion.

Besides distortion, other aberrations vary with the pupil coordinates. Fig. 3-8 shows the intermediate image seen by the surface subareas in the presence of spherical aberration, coma, and astigmatism, and Fig. 3-7 shows the distorted detector grid on the intermediate image plane. The blue and green arrows in Fig. 3-8(a) indicate the intermediate images of the detector seen by an on-axis point O and off-axis points E and F in the presence of spherical aberration, while Fig. 3-7(a) shows the distorted detector grid seen by point F . From Fig. 3-8(a) and Fig. 3-7(a) we can see that spherical aberration shifts the pixel centers by a constant distance due to the fact that primary spherical aberration is field-independent. In contrast to spherical aberration, the impact of coma and astigmatism is field-dependent and varies for the subareas in the tangential and sagittal directions. As shown in Fig. 3-8(b), the three arrows $O'P$, $O'S$, and $O'T$ indicate the intermediate image

seen by on-axis point O , sagittal points C and D , and tangential points E and F in the presence of coma, and the distorted detector grids seen by the tangential point E and sagittal point C are shown in Fig. 3-7(b, c), from which we can see that coma distorts and enlarges the detector grid. While both coma and spherical aberration do not bend the planar intermediate images, astigmatism and Petzval curvature do. As shown in Fig. 3-8(c), the red and green arrows indicate the curved intermediate image seen by the tangential points E and F , and sagittal points C and D in the presence of astigmatism, and we can see that the radii of curvature and sizes of the curved intermediate images are different for the tangential and sagittal points, so the intermediate images seen by the subareas have toroidal shapes. Fig. 3-7(d, e) show the distorted detector grid in the presence of astigmatism and field curvature.

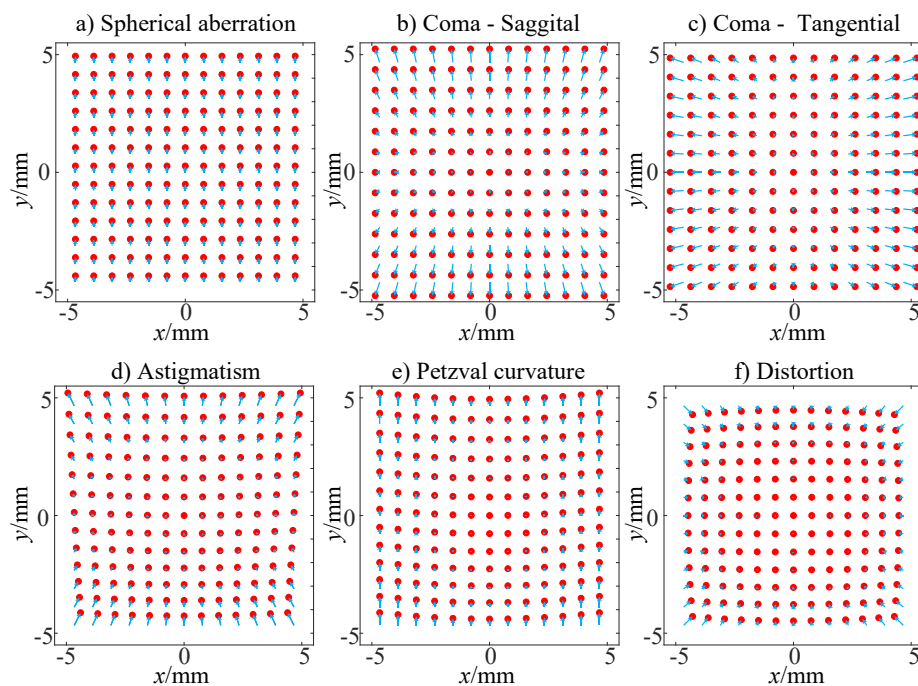


Fig. 3-7. Distorted detector grid due to aberrations seen by a point on the optical surface. The red dots indicate the locations of pixel centers in the distorted intermediate image while the blue arrows indicate the shift of pixel centers due to aberrations. The wave aberrations are represented by Seidel coefficients. a) shows spherical aberration with $S_I = -0.1$ mm; b,c) show coma with $S_{II} = -0.1$ mm; d) shows astigmatism with $S_{III} = -0.1$ mm; e) shows Petzval curvature with $S_{IV} = -0.2$ mm; f) shows distortion with $S_V = 0.2$ mm. c) corresponds to a sagittal point on the optical surface while the rest correspond to tangential points. The size of the paraxial intermediate image is $4.9 \text{ mm} \times 4.9 \text{ mm}$.

The impact of aberrations on the intermediate image has usually been ignored in previous approaches based on inverse raytracing [12], but in reality the impact can be too large to be neglected because the shift of pixel positions due to aberrations induces errors in the calculation of the solid angles subtended by pixels as well as the values of the BRDF in the scattering directions. In general, if the shift of pixel position is larger than the width of a pixel, the impact of aberrations should be considered.

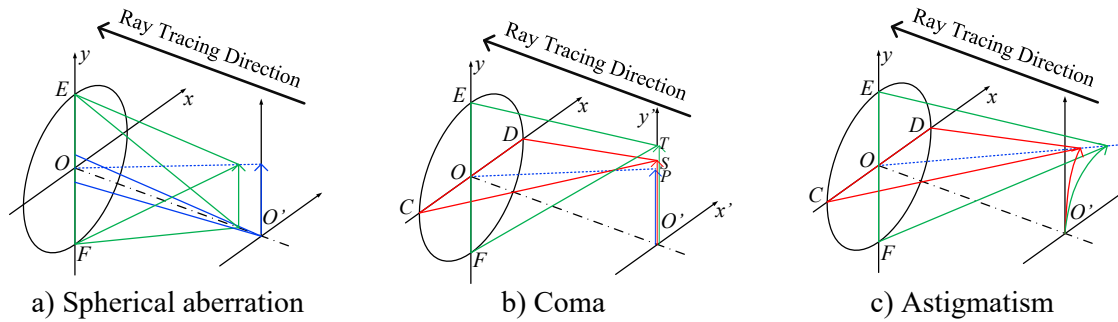


Fig. 3-8. Impact of aberrations on the intermediate image of the detector. The arrows indicate the intermediate images seen by different points on the optical surface.

3.3.2.2 Inverse raytracing method

The above-mentioned method to calculate the acceptance angle based on the detector intermediate image can be applied to circularly symmetric systems without large high order aberration contribution from its elements. However, for systems without circular symmetry, or if high order aberrations are of concern, the intermediate image method may be difficult to implement, or the efficiency can be greatly decreased.

A more general method to calculate the acceptance angle is based on inverse real raytracing. In this method, rays are traced from several points (nodes) on the detector into the optical system, as indicated by the blue arrows in Fig. 3-9. The ray density of the reversely traced ray cones must be high enough, so that every subarea on the optical surfaces intersects with at least one ray from each node.

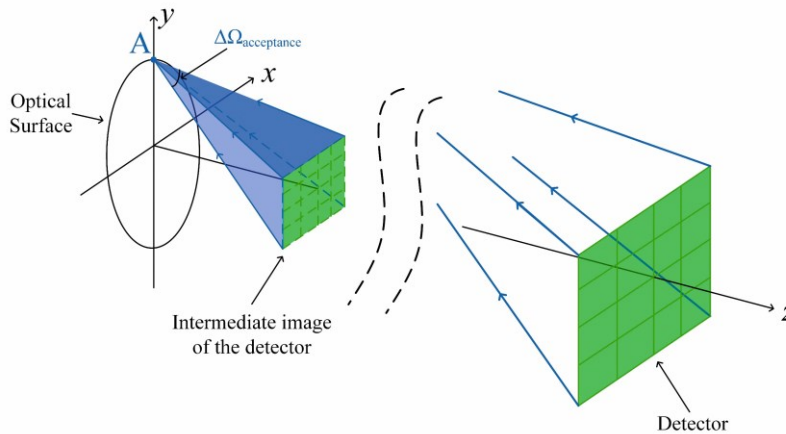


Fig. 3-9. Schematic illustration of the inverse raytracing and hybrid methods. The dashed grid in the detector intermediate image shows to the discretization of the detector intermediate image in the hybrid method.

As shown in Fig. 3-9, the detector acceptance angle $\Delta\Omega_{\text{acceptance}}$ in subarea A is determined by the cone angle of the blue pentahedron formed by the reversely traced rays. Compared to the intermediate image method, the inverse raytracing method is more robust, but sacrifices efficiency since rays need to be traced from all corners of the pixels in order to calculate the irradiance distribution on the detector. Therefore, the inverse raytracing

method can only be used to calculate the total scattered flux rather than the irradiance distribution on the detector.

3.3.2.3 Hybrid method

In order to alleviate the limitation of the inverse raytracing method in the calculation of the irradiance distribution, a hybrid method which combines the inverse raytracing method with the intermediate image method is proposed. As shown in Fig. 3-9, in the hybrid method, rather than calculating the total acceptance angle of the detector, the acceptance angles of the pixels are calculated by discretizing the detector intermediate image on a mesh grid, which is obtained by applying perspective transformations on the ideal grid of the detector intermediate image. The perspective transformation matrix can be calculated conveniently once the coordinates of the four detector corners in the intermediate image plane are determined by inverse raytracing.

The relationship between the hybrid method with the inverse ray tracing and intermediate image methods is depicted in Fig. 3-10. From the flowchart in Fig. 3-10 we can see that the first step of the hybrid method is the same as the inverse ray tracing method, in which we inversely trace several ray bundles from the center and corners of the detector to the optical surfaces. Based on the inverse ray tracing result, we calculate a linear transformation matrix of the detector grid based on the intermediate image locations of the center and corners of the detector seen by the subareas of the optical surface. Subsequently, the linear transformation matrix is applied to distort the detector grid, based on which the subarea-pixel coupling algorithm of the intermediate image method is applied to calculate the contribution of each surface subarea to the straylight intensity in each pixel.

In Fig. 3-10, the blue, red, and green colors indicate the simulation steps implemented in the intermediate image, inverse raytracing, and hybrid methods. From the flowcharts we can see that the main difference between the three methods is the calculation of the detector intermediate image and the coupling between the subareas on the optical surface and the detector. Furthermore, we can also see that the intermediate image method is different from the inverse raytracing method in every step except for the calculation of the irradiance on the optical surface, while the hybrid method inherits inverse real raytracing from the inverse raytracing method and the coupling between the subareas and pixels from the intermediate image method.

The hybrid method is able to effectively calculate the straylight distribution on the detector for systems with arbitrary geometry, but due to the application of linear transformations to obtain the distorted detector grid, only the linear dependence on the field coordinate such as tilt, linear coma, or field-independent spherical aberration can be accurately modelled. The impact of aberrations with higher order dependence on the field coordinate are not accurately described, this being the main limitation of the accuracy and

robustness of the hybrid method. Further details about the hybrid method will be discussed in Sec. 5.1.

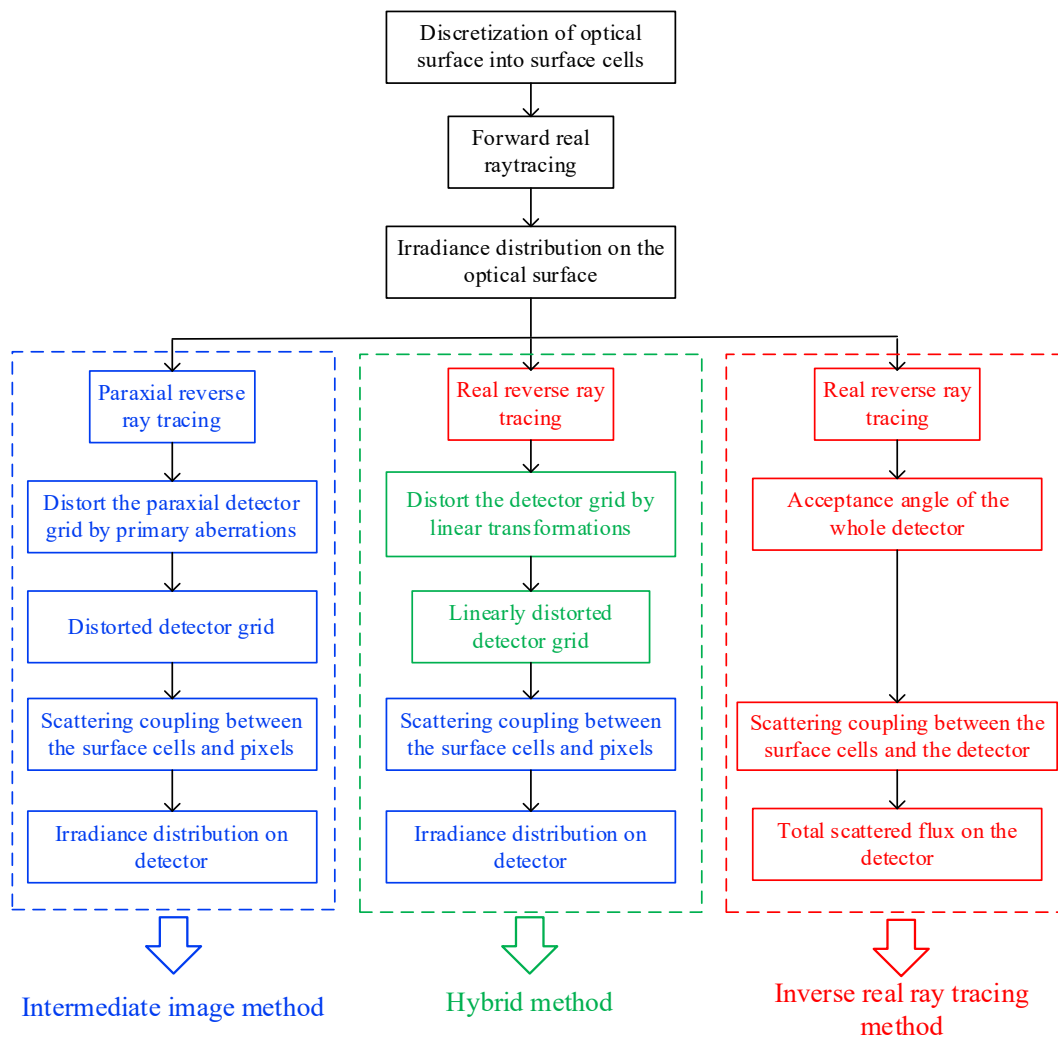


Fig. 3-10. Flowchart of the three methods to implement the phase space model.

3.4 Geometrical volume scattering simulation

For volume scattering, the incident light can be scattered anywhere in the scattering volume rather than only on the surfaces. Therefore, in order to apply the phase space method in the simulation of volume scattering, we have to create artificial surfaces in the scattering volumes, on which the phase space coupling methods described in the last sections can be applied. As shown in Fig. 3-11, the artificial scattering surfaces are created by slicing the scattering volume along the optical axis. Each of the z-slices obtained after this process is considered as a scattering surface, whose TIS is determined by the thickness of the z-slice and the scattering coefficient of the medium, while the scattering distribution function is determined by the scattering phase function of the scatterer in the volume.

For each z-slice, the phase space overlapping of the scattered light and detector acceptance gives the straylight contribution from this z-slice. Finally, the total straylight contribution from an element is obtained by summing up the contribution from each z-slice of this element. It should be noted that, similar to surface scattering simulation, here only single scattering can be considered since the determination of the illumination area and acceptance angles requires deterministic raytracing, meaning that the light scattered within a z-slice is propagated to the image space without secondary scattering. Therefore, the phase space method can only be applied for the simulation of weakly-scattering medium such as transparent glass or air gap, while it is not applicable if turbid medium is involved.

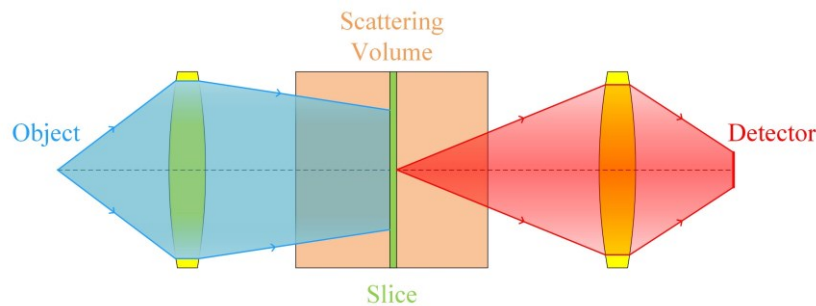


Fig. 3-11. Schematic sketch of the simulation of volume scattering by slicing the optical system along the optical axis. The blue ray cone is the illumination ray cone, and the red ray cone is the acceptance ray cone of the detector for the central subarea of the z-slice.

Furthermore, the thickness of the slice, or the number of slices in an optical element has a large impact on the accuracy and efficiency of the simulation. Since the phase space coupling should be repeated for every slice, the calculation time is linearly dependent on the number of z-slices. On the other hand, the impact of slice density on the accuracy depends on the shape of the element and the NA of the beam in the element. For example, denser z-slices are required in the elements with strongly curved surfaces to describe the exact shape of the surfaces, and elements in which the beam NA is large also require thin slices to describe the fast-changing beam size.

3.5 Field windows

In the geometrical phase space models, special care must be taken with the optical elements where the ray bundles are strongly separated, because the subareas of such surfaces or slices cannot see the entire intermediate image due to the truncation of light by the apertures downstream from the optical element. An example of this special case is shown in Fig. 3-12, which shows a schematic sketch of a retrofocus lens with inversely traced ray trajectories. Here we can see that the inversely traced off-axis ray bundle is vignetted by two elements and is strongly separated from the on-axis ray bundle on the front element L1. Consequently, the surface subareas on L1 are only able to see a part of the detector through the apertures downstream from L1. The intermediate images of these

apertures act as field windows through which the surface subareas on L1 see the detector. As shown in Fig. 3-12, the field windows for the surface subareas on L1 are the intermediate images of L2, L3, and the stop, which coincide with themselves if we neglect the small focal power of L2. For the two points O and B on L1, the parts of the detector that the two points can see are determined by the overlap of the projections of all field windows on the detector intermediate image plane through O and B , as indicated by the orange and pink arrows in Fig. 3-12. Therefore, for such surfaces displaying ray bundle separation, the field windows must be considered in the calculation of the solid acceptance angle subtended by the detector.

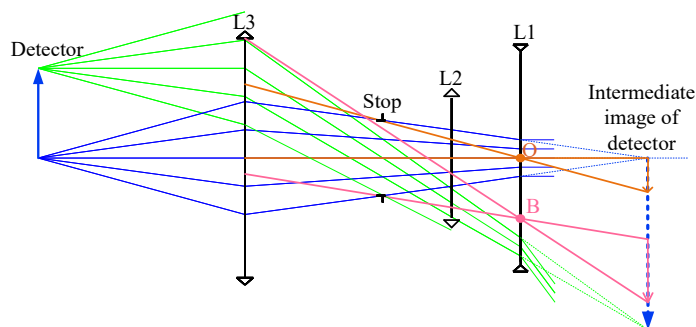


Fig. 3-12. Schematic sketch of a retrofocus lens in a reversed order. The blue dotted arrow shows the intermediate image of the detector seen by L1. Points O and B are an on-axis point and an off-axis point on L1, and the orange and pink arrows indicate the portions of the intermediate image seen by these two points.

3.6 Physical optics simulation based on the Wigner function

Although the phase space methods described in the previous sections shows great advantage in accuracy and efficiency of light scattering simulation, they are limited by the fact that they are completely ray-based. Therefore, the physical aspects of light such as partial coherence, diffraction and interference cannot be modelled, which prohibits us from calculating the PSFs of real optical systems. Additionally, the requirement for deterministic raytracing makes it impossible to model multiple scattering with the phase space methods.

To solve these problems, we propose a Wigner function-based approach to propagate partially coherent light in real optical systems with scattering surfaces. As mentioned in Sec. 2.5, The advantages of the Wigner function in the modelling of light scattering is its capability to characterize and propagate partially coherent light field and to include diffraction effects, which is normally seen in reality. As we will see later, the mid and high spatial frequency components of the PSD have different influence on the spatial coherence of light and Wigner function is able to tackle both of them in a unified model. Another advantage of the Wigner function is its direct connection with the angular distribution of light, making it possible for the Wigner function to be directly coupled

with the BSDF. In this section, we first introduce a definition of light coherence based on both time and ensemble average, and then the scattering of light by different types of surface errors can be interpreted by the coherence theory.

3.6.1 Ensemble average coherence

According to the discussion in Sec. 2.2.1, both statistical modelling and exact field propagation are necessary to analyze the impact of scattering from different spatial frequency components of the residual surface errors. In order to combine them in a unified model for real optical systems, we need to consider light as partially coherent. Therefore, it is necessary to clarify the definition of spatial coherence based on time average and ensemble average. Here we restrict our discussion to the one-dimensional case for the sake of simplicity and without loss of generality.

The classical definition of spatial coherence is based on the time average [18], with the mutual intensity defined as

$$\mathbf{J}(P_1, P_2) = \langle \mathbf{u}(P_1, t) \mathbf{u}^*(P_2, t) \rangle, \quad (3.13)$$

where $\mathbf{u}(P, t)$ is the field at point P and time t , and the brackets denote the time average. According to this definition, the light field from a monochromatic point source remains fully coherent throughout an optical system if the optical surfaces are stationary. However, this is not helpful for the interpretation of light scattering from the perspective of spatial coherence because scattering has no impact on the coherence state of light under this time-average definition. Therefore, it makes sense to utilize the definition of spatial coherence based on the ensemble average, in which the mutual intensity should be calculated as an average over an ensemble of different scattering surfaces denoted by $E[\]$

$$\bar{\mathbf{J}}(P_1, P_2) = E[\mathbf{u}(P_1, t) \mathbf{u}^*(P_2, t)]. \quad (3.14)$$

If we consider a fully coherent light field incident perpendicularly on a thin transmissive scattering surface, the incident light field at a point P before the surface can be written as

$$\mathbf{u}_i(P, t) = \mathbf{A}(P) \exp(-j2\pi vt), \quad (3.15)$$

while the mutual intensity function according to Eq. (3.14) is

$$\bar{\mathbf{J}}_i(P_1, P_2) = \mathbf{A}(P_1) \mathbf{A}^*(P_2). \quad (3.16)$$

Assuming that the surface is non-absorbing and applying the thin element approximation (TEA), we can write the field after the surface as

$$\mathbf{u}_s(P, t) = \mathbf{A}(P) \exp(-j2\pi vt) \exp(-j\varphi_p), \quad (3.17)$$

where φ_p is the random phase delay induced by the scattering surface at point P .

Substituting Eq. (3.17) into Eq. (3.14), we obtain the mutual intensity of the light after the scattering surface

$$\bar{\mathbf{J}}_s(P_1, P_2) = \mathbf{A}(P_1) \mathbf{A}^*(P_2) E\{\exp[-j(\varphi_1 - \varphi_2)]\}. \quad (3.18)$$

The phase variations φ_1 and φ_2 can be considered as two random variables which are proportional to the surface height of the scattering surface at P_1 and P_2 , and the mathematical expectation $E\{\exp[-j(\varphi_1 - \varphi_2)]\}$ can be calculated by the joint characteristic function $\Phi(\omega_1, \omega_2)$ of the random variables [14]

$$E\{\exp[-j(\varphi_1 - \varphi_2)]\} = \Phi(1, -1). \quad (3.19)$$

If φ_1 and φ_2 are zero-mean Gaussian processes and jointly normal, the joint characteristic function can be analytically calculated. Substituting Eqs. (3.16) and (3.19) into Eq. (3.18), we get

$$\bar{\mathbf{J}}_s(P_1, P_2) = \bar{\mathbf{J}}_i(P_1, P_2) \exp[-\sigma_\varphi^2(1 - C_{12} / \sigma_\varphi^2)], \quad (3.20)$$

where σ_φ^2 is the deviation of the phase variation induced by the scattering surface and C_{12} is the autocovariance function of the phase variations at P_1 and P_2 . If we extract the spatially variant part, we can rewrite the mutual intensity function as

$$\begin{aligned} \bar{\mathbf{J}}_s(P_1, P_2) &= \bar{\mathbf{J}}_i(P_1, P_2) \left[\exp(-\sigma_\varphi^2) + \frac{1 - \exp(-\sigma_\varphi^2)}{\exp(\sigma_\varphi^2) - 1} [\exp(C_{12}) - 1] \right] \\ &= \bar{\mathbf{J}}_i(P_1, P_2) \cdot (1 - TIS) + \bar{\mathbf{J}}_i(P_1, P_2) \cdot TIS \cdot \frac{\exp(C_{12}) - 1}{\exp(\sigma_\varphi^2) - 1}, \end{aligned} \quad (3.21)$$

where $TIS = 1 - \exp(-\sigma_\varphi^2)$ is the total integrated scattering and $1 - TIS$ is the ratio of energy in the specular transmission. Eq. (3.21) provides useful insights into the impact of the scattering surface on the spatial coherence of the incident light. We can see that the mutual intensity function after the scattering surface has two mutually incoherent components. One is the specularly transmitted part which has the same degree of spatial coherence as the incident light with attenuated intensity, while the other part is the diffusively transmitted part, whose spatial coherence is determined by the degree of coherence of the incident light together with the statistical property of the scattering surface. The percentage of the energy contained in each component is determined by the TIS, which depends on the roughness of the scattering surface. The above discussion assumes coherent incident light, but the conclusion can be easily extended to partially coherent incident light by recognizing that the mutual intensity function of partially coherent light can be represented by the sum of the mutual intensity functions of a set of fully coherent but mutually incoherent light fields [19].

Furthermore, if the scattering surface is optically smooth, which means that the surface height variation of the scattering surface is much smaller than one wavelength ($\sigma_\varphi \ll 2\pi$, $C_{12} \ll 4\pi^2$), we can further simplify Eq. (3.21) by expanding the exponential

terms into a Taylor series and retaining only the first two terms. Then the mutual intensity function for smooth surfaces can be written as

$$\bar{\mathbf{J}}_s(P_1, P_2) = \bar{\mathbf{J}}_i(P_1, P_2) \cdot (1 - TIS) + \bar{\mathbf{J}}_i(P_1, P_2) TIS \cdot \rho(P_1, P_2), \quad (3.22)$$

where $\rho(P_1, P_2) = C_{12} / \sigma_\phi^2$ is the autocorrelation coefficient of the phase variations.

From Eq. (3.22) we learn that for a coherent incident light field, the scattered part of the field after a surface shares the same autocorrelation coefficient as the profile of that surface. For high frequency microroughness, their correlation length is usually quite small, and the scattered component can be considered to have very low spatial coherence, meaning that the scattered light is only partially coherent within very small areas of coherence defined by the correlation length of the scattering surface. Therefore, if the irradiance distribution on a plane after the scattering surface is to be calculated, the contribution from the specular component can be obtained by scaling the amplitude of the incident field by $\sqrt{1 - TIS}$ and then propagating it to the target plane. Separately, the contribution from the scattered component can be obtained by incoherent superposition of the light scattered from each area of coherence. Therefore, Eq. (3.22) serves as the theoretical foundation for the ray-based method to simulate light scattering in optical systems, which are commonly used despite the fact that their theoretical background is rarely discussed. The ray-based method traces rays to the scattering surfaces and splits the incident rays into specular and scattered components after incidence. The specular part undergoes reflection or refraction, while the scattered part changes its propagation direction according to the BSDF of the surface. Finally, the irradiance distribution of the scattered component is obtained by incoherently superposing all scattered rays, which is justified by the fact that the light scattered by different correlation areas of the surface is mutually incoherent. Therefore, the preassumption of the ray-based method is that the correlation areas on the scattering surface should be much smaller than the size of the scattering surface, such that the light scattered by different areas of the surface can be incoherently superposed.

This assumption is usually fulfilled for high frequency surface roughness, but fails for MSF errors because of the large correlation length and because the wide-sense stationarity is no longer guaranteed. Therefore, the statistical characterization of the scattering property of surfaces using the BSDF can only be applied to surface roughness. For MSF errors, similar to what we have concluded in the last section based on the PSF and MTF, the finite degree of spatial coherence of the scattered field requires us to take the interference and diffraction effect into account. Consequently, rigorous beam propagation methods based on the exact metrology data of the MSF structures need to be applied to evaluate their scattering effect.

Therefore, under the ensemble average-based definition of light coherence, scattering of light by surfaces can be interpreted as splitting the incident light field into specular and scattered parts which are mutually incoherent with each other, resulting in a reduction of coherence. This enables us to model the scattered light as partially coherent and to use both statistical and analytical models simultaneously to simulate light scattering due to residual surface errors.

3.6.2 Propagation and scattering of partially coherent beams based on the Wigner function

According to the previous discussion, the capability of the Wigner function to propagate partially coherent light enables us to model multiple scattering from an arbitrary number of surfaces. Furthermore, by combining the statistical and analytical surface scattering models, the Wigner function can also model light scattering from different frequency components of the PSD simultaneously. In this section, we demonstrate the method to simulate surface scattering in optical systems by the propagation of partially coherent light based on the Wigner function. For simplicity, we restrict our discussion to 2D Wigner functions.

Applying the time average-based and ensemble average-based definitions of light coherence simultaneously and recall the definition of the cross-spectral density function in Eqs. (2.31), we can redefine the cross-spectral density function by time and ensemble averages

$$\Gamma(x, \Delta x) = E \left\{ \left\langle E \left(x + \frac{\Delta x}{2} \right) E^* \left(x - \frac{\Delta x}{2} \right) \right\rangle \right\}, \quad (3.23)$$

where the $E\{\langle \rangle\}$ denotes time and ensemble averages.

Consider an incident light field E_i with cross-spectral density function Γ_i propagating through a thin scattering surface. Applying the TEA, we can write the cross-spectral density function of the scattered light field as

$$\begin{aligned} \Gamma_s(x, \Delta x) &= E \left\{ \left\langle E_i \left(x + \frac{\Delta x}{2} \right) E_i^* \left(x - \frac{\Delta x}{2} \right) t \left(x + \frac{\Delta x}{2} \right) t^* \left(x - \frac{\Delta x}{2} \right) \right\rangle \right\} \\ &= \left\langle E_i \left(x + \frac{\Delta x}{2} \right) E_i^* \left(x - \frac{\Delta x}{2} \right) \right\rangle \cdot E \left\{ t \left(x + \frac{\Delta x}{2} \right) t^* \left(x - \frac{\Delta x}{2} \right) \right\} \\ &= \Gamma_i(x, \Delta x) \cdot \Gamma_t(x, \Delta x), \end{aligned} \quad (3.24)$$

where Γ_i and Γ_t are the cross-spectral density functions of the incident field and the scattering surface, and the transmission function $t(x)$ is the modulation of the scattering surface on the incident field. For a non-absorbing thin element, $t(x)$ can be written as

$$t(x) = \exp \left[j2\pi(n_2 \cos \theta_s - n_1 \cos \theta_i)h(x) \right], \quad (3.25)$$

where $h(x)$ is the height of the surface in waves, θ_i and θ_s are the incident and scattering angles, n_1 and n_2 are the refractive indices of the medium before and after the scattering surface.

Substituting Eq. (3.24) into Eq. (2.32), and applying the convolution theorem of the Fourier transform, the Wigner function of the scattered field can be calculated as

$$W_s(x, u) = W_i(x, u) \otimes W_t(x, u), \quad (3.26)$$

where $W_i(x, u)$ and $W_t(x, u)$ are the Wigner functions of the incident field and the scattering surface. Therefore, the Wigner function of the light field after a scattering surface can be calculated as the convolution of the incident Wigner function with the Wigner function of the surface transmission function in the angular domain. Next, we will apply the above derivation and demonstrate the modelling of surface scattering from different types of surface errors.

3.6.2.1 Scattering from high spatial frequency (HSF) micro roughness

According to the discussion in Sec. 3.6.1, the surface height variations of HSF errors can be modelled by Gaussian processes with wide sense stationarity. Based on this assumption and applying the paraxial approximation, we substitute Eq. (3.25) into Eqs. (2.32) and (3.23), then follow the derivations from Eqs. (3.18) - (3.21) to obtain the Wigner function of the scattering surface

$$W_t(x, u) = (1 - \text{TIS})\delta(u) + S(u). \quad (3.27)$$

Here $\delta(u)$ is the delta function, and $S(u)$ is the angle spread function which is linearly dependent on the PSD of the scattering surface according to the Harvey-Shack scattering model [4]

$$S(u) = \frac{4\pi^2(n-1)^2}{\lambda^4} \text{PSD}\left(\frac{u}{\lambda}\right), \quad (3.28)$$

where n is the refractive index of the thin element, and λ is the wavelength of the incident light.

Subsequently, by substituting Eq. (3.27) into Eq.(3.26), we can calculate the Wigner function of the light field after the scattering surface

$$W_s(x, u) = (1 - \text{TIS}) \cdot W_i(x, u) + W_i(x, u) \otimes S(u). \quad (3.29)$$

Therefore, similar to the mutual intensity function, the Wigner function after the scattering surface is composed of two parts. The first is the specular transmission which

can be calculated by scaling the incident Wigner function by (1-TIS), while the second is the scattered portion which is calculated by convolving the incident Wigner function with the angle spread function $S(u)$. It should be noted that here $S(u)$ is independent of the incident angles because the derivation of Eq. (3.29) is based on the paraxial approximation and the TEA, while in reality, if the paraxial approximation is violated, the transmission function $t(x)$ is also dependent on the incident angle, as is the angle spread function. Therefore, the angle spread function $S(u)$ in Eq. (3.29) is only a slice of the BSDF for normal incidence, while for general cases the Wigner function of the scattered light should be calculated by an integration of the BSDF over the incident angles

$$W_s(x, u) = (1 - TIS) \cdot W_i(x, u) + \int_{-\pi/2}^{\pi/2} W_i(x, u_i) BSDF(u_i, u) du_i. \quad (3.30)$$

In cases where the paraxial approximation is fulfilled or if the BRDF is shift-invariant for different incident angles, Eq. (3.30) reduces to Eq. (3.29). As an example, Fig. 3-13 shows the propagation of a top-hat Gaussian Schell beam through a non-absorbing scattering window with HSF errors. From the figure we can see that the surface scatters part of the incident beam to large angles, and consequently the spatial extent of the beam increases with propagation in z .

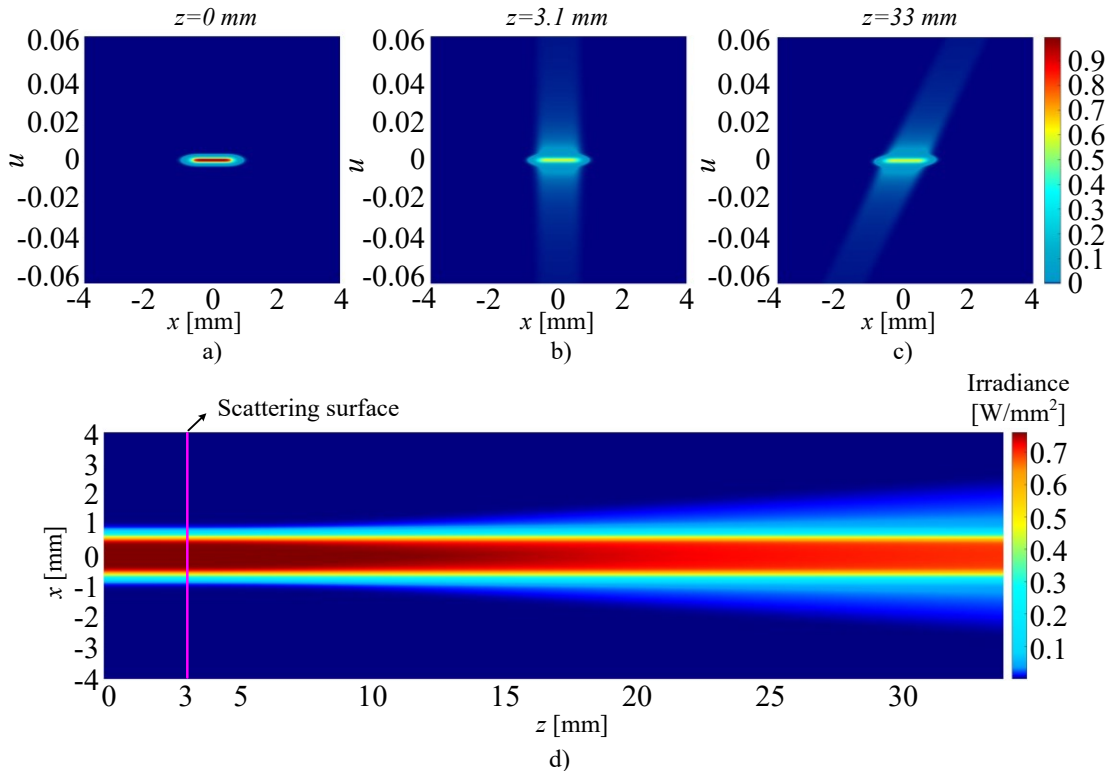


Fig. 3-13. Propagation of a Gaussian Schell beam through a scattering surface at $z = 3$ mm with HSF surface errors. The first three figures show the Wigner function of the beam before the scattering surface (a), after the scattering surface (b), and after 30 mm of free space propagation (c). d) shows the irradiance distribution of the beam.

During simulation, the integration in Eq. (3.30) are evaluated at every x -location, which can be done efficiently by employing vectorization of the calculations. The calculation time for the results in Fig. 3-13 is 75 s using an Intel Core i5 CPU (6 cores @ 3.3 GHz), the resolution of the Wigner function is $2048(x) \times 512(u)$, and the total number of z -steps is 90. The BSDF of the scattering surface is modelled by the generalized Harvey-Shack (GHS) theory [4] and the PSD of the surface is modelled by the K-correlation method [12]. The wavelength of the monochromatic beam is $1 \mu\text{m}$, the correlation length of the surface error is $50 \mu\text{m}$, the refractive index of the scattering surface is 1.55, and the standard deviation of the surface error is $0.2 \mu\text{m}$.

3.6.2.2 Scattering from mid-spatial frequency (MSF) surface errors

MSF errors are usually related to the specific manufacturing method and possess intrinsic patterns. Therefore, statistical methods are no longer ideal for modelling such errors, and the scattering effect should be modelled using the surface transmission function $t(x)$ based on exact metrology data.

As an example, we consider a surface with concentric grooves, which is a typical structure resulting from diamond turning. In the 1D case, along the radial direction, the concentric grooves can be modelled by a piecewise-parabolic function, but in reality, the MSF surface structure is not perfectly periodic due to the vibration of the tool tip [6]. In most real cases, these vibrations are the dominant reason for MSF. The vibrations in the thrust and feed directions induce displacement of the parabolic function in the vertical and horizontal directions. Here we assume that these random displacements follow Gaussian distributions, based on which we have generated an example of the MSF structures as shown in Fig. 3-14. The period and the PV value of the MSF structure in Fig. 3-14 are $T = 40 \mu\text{m}$ and $h_{pv} = 0.5 \mu\text{m}$, and the standard deviations of the tool tip displacement in the thrust and feed directions are $\sigma_t = 0.05 \mu\text{m}$ and $\sigma_f = 4 \mu\text{m}$.

Based on the MSF structure stated above, we have simulated the paraxial propagation of a Gaussian Schell beam through the MSF structure and a single lens based on Eqs. (3.24)-(3.25), and the results are shown in Fig. 3-15 and Fig. 3-16. From Fig. 3-15(a) we can see that the MSF structures transform the top-hat beam into many beamlets. In the focal plane of the single lens, we observe strong side lobes on both sides of the central peak. Fig. 3-15(b) shows the PSFs of the system for four MSF structures with different periods and PV values. By comparing the four PSFs, we can see that the PV values of the MSF structures govern the intensity of the side lobes, with larger PV values resulting in weaker central peaks and stronger side lobes. On the other hand, the scattering angles are dependent on the periods of the MSF structures, with larger periods corresponding to smaller scattering angles.

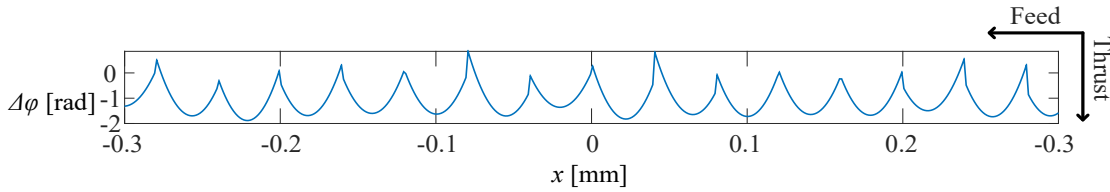


Fig. 3-14. An example of surface structure of MSF surface error. Displacement of the tool tip due to vibrations in thrust and feed directions follows Gaussian distributions, and phase delay $\Delta\phi$ is calculated for the wavelength of $1 \mu\text{m}$.

The blue curve in Fig. 3-15(b) corresponds to a critical frequency at which the valleys between the side lobes and central peak disappear and the central peak is broadened, resulting in resolution reduction of the optical system. Therefore, the critical frequency can be used to estimate the impact of MSF structures on image quality. For MSF structures below this critical frequency, the major impact is resolution reduction, while for MSF structures above the critical frequency, the resolution is not significantly reduced, but stray light structures such as ghost images or halos around bright objects will be observed due to the strong side lobes.

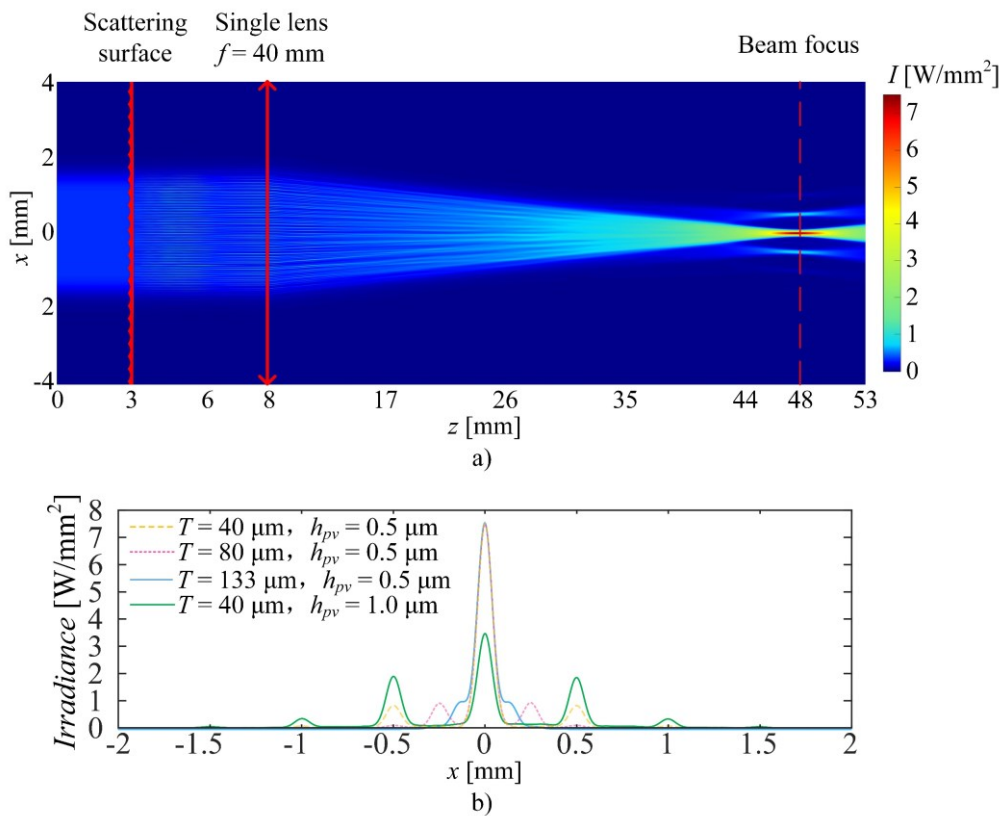


Fig. 3-15. Propagation of a Gaussian Schell beam through a scattering surface with MSF structures and a single lens. a) shows the irradiance distribution on different z -planes for the scattering surface described in Fig. 3-14, the single lens having a focal length of $f = 40 \text{ mm}$. b) shows the x -cross-section of the irradiance distribution at the beam focus for four MSF structures with different periods and PV values.

Fig. 3-16 shows the Wigner functions of the beam through propagation. In Fig. 3-16(b, d) we observe negative values of the Wigner function, which is a consequence of phase space destructive interference [20]. Finally, the intensity distribution at the focal plane of

the single lens results from the interference of all the beamlets formed by the piecewise-parabolic structures. The runtime of the simulation is 16 minutes for 130 z -steps and a Wigner function resolution of 2048×2048 using the same computer platform as in the last section. The major limitation for the runtime is the complicated angular distribution of the Wigner function after the scattering surface, which precludes the use of coarser sampling in the angular domain as in the last section. However, the processing time can be reduced to 7 seconds if we only calculate the irradiance distribution at the focal plane, since intermediate results are not usually needed.

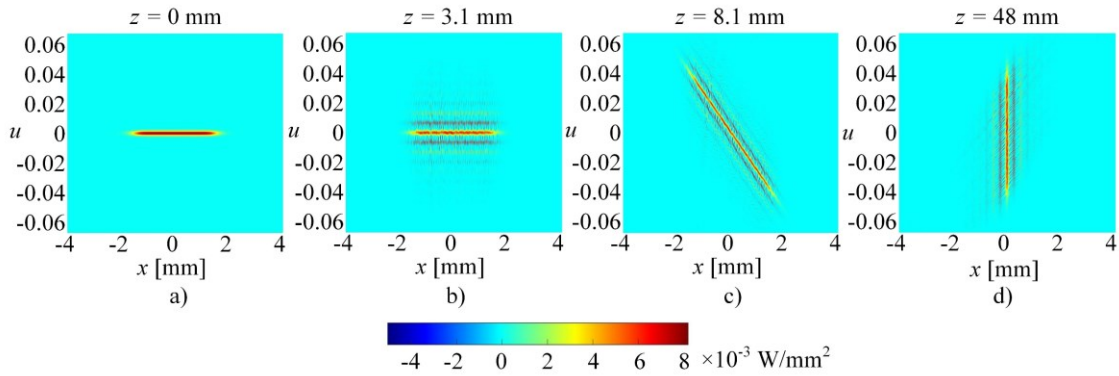


Fig. 3-16. Wigner functions of the Gaussian Schell beam propagated through the optical system described in Fig. 3-15(a) for an MSF structure with $T = 40 \mu\text{m}$ and $h_{pv} = 0.5 \mu\text{m}$.

4 Realization of Simulation Algorithms

In this chapter, the realization of the geometrical phase space methods is demonstrated. As an example, the geometrical phase space model is implemented in the simulation of the autofluorescence effect of microscope lenses and scattering effect in plastic lenses.

Based on the results of the autofluorescence simulation, the errors of the phase space method for volume scattering are analyzed. After the origins of the errors are identified, methods are proposed to improve the accuracy of the simulation. Furthermore, due to the high efficiency and automation of the phase space method, we are able to calculate the autofluorescence effect of a large collection of microscope lenses and perform a systematic analysis of the autofluorescence performance of different types of microscope lenses. Based on the systematic analysis, we can define the critical parameters of the objectives and lens group structures which result in large autofluorescence contribution. Additionally, the design strategies of microscope lenses with low autofluorescence level are discussed based on the systematic analysis.

4.1 Volume scattering

The autofluorescence effect of optical glasses and the scattering of plastic materials have large impact on the image quality of microscope objectives and plastic lenses. These two types of straylight can be modelled as volume scattering because they are generated inside the optical material. In the conventional Monte Carlo methods that model volume scattering, a large number of rays are launched into the optical system from the light source, whose area and NA are sampled with random rays. The rays are then propagated through the optical system, in which they are scattered at random locations determined by the mean free paths of the scattering medium. Finally, the rays that hit the detector plane is collected and the weightings of the rays in every pixel of the detector are counted. To model the process of fluorescent-light generation, the Stokes shift of the wavelength is taken into account as well as the isotropic radiation of the molecular re-emission. Furthermore, the specular and scattered rays are attenuated during the propagation according to the Beer-Lambert law. The problem with this kind of calculation scheme is the poor efficiency due to the large number of random rays that have to be traced to accurately sample the light source and the angular distribution of the scattered light. Insufficient sampling of the light source or the scattering distribution function result in signals with small statistical noise. The corresponding long computational time prevents this simple brute-force method to be used in system optimization. In contrary, due to the absence of Monte Carlo raytracing, the phase space method shows much higher efficiency, and since the results of the phase space method is calculated quasi-analytically, the accuracy of the phase space method is also advantageous compared to the Monte Carlo method.

4.1.1 Description of the method

Considering that the autofluorescence is generated everywhere in the lens volume, it can be modelled by volume scattering with a wavelength shift, which can be efficiently simulated with the phase space method.

As discussed in Sec. 3.4, the first step of autofluorescence simulation is to slice the microscope lenses along the optical axis. Fig. 4-1 shows a schematic sketch of the autofluorescence generation and acceptance in a slice with a thickness of Δz_j , which is chosen to be small such that the height of the ray bundle can be considered to be constant within the z-slice. Furthermore, we consider transmission illumination and only one illumination point on axis for simplicity. Fig. 4-2 shows the corresponding phase space description of the generation and acceptance of the fluorescence light in this z-slice.

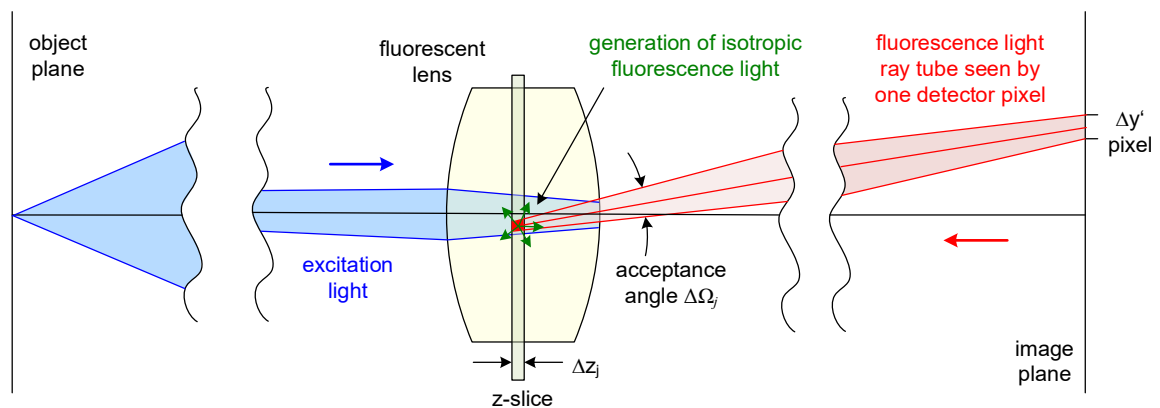


Fig. 4-1. Schematic setup of the model with excitation light coming from the object plane, total fluorescent light created in the lens and detected fluorescence light reaching one detector pixel.

In the lens, the intensity of the excitation is evaluated by photometric formula with the help of the ray data. Since the direction of the fluorescence light is completely independent on the direction of the excitation light, the fluorescent light is generated isotropically into 4π solid angle. Additionally, we also assume that there is no apodization in the system and the excitation light is uniformly distributed in the spatial domain, this assumption is not always true, especially in elements in which the aperture angles of the ray bundles are large. The error induced by this assumption and the corresponding compensation methods are discussed in detail in the next section. Based on the above assumptions, as indicated by the blue area in Fig. 4-2, the phase space distribution of the fluorescence light is uniform, and we only need to trace two marginal rays to determine the spatial boundaries of the phase space area. For the acceptance of the fluorescence light, only those rays coming out of the illuminated area in the acceptance solid angle can reach the pixel with size $\Delta y'$ on the detector. The corresponding light pipe is sketched in red in Fig. 4-1 and Fig. 4-2. In Fig. 4-2, the spatial boundaries of the acceptance area are determined by the spatial distance Δy between two coma rays which are traced from the center of the pixel to the z-slice. On the other hand, the angular boundaries of the

acceptance area are determined by aiming two rays from the upper and lower edges of the pixel towards the center of the spatial acceptance area, and the angular difference Δu between these two rays gives the angular range of the phase space acceptance area. Since real ray aiming is rather time-consuming and the acceptance angles in microscope lenses are usually rather small, we aim these two rays paraxially by the ABCD matrix, which allows us to determine the initial directions of the aimed rays by solving linear equations and therefore is extremely efficient. The accuracy of the paraxial ray aiming method is discussed in the next section.

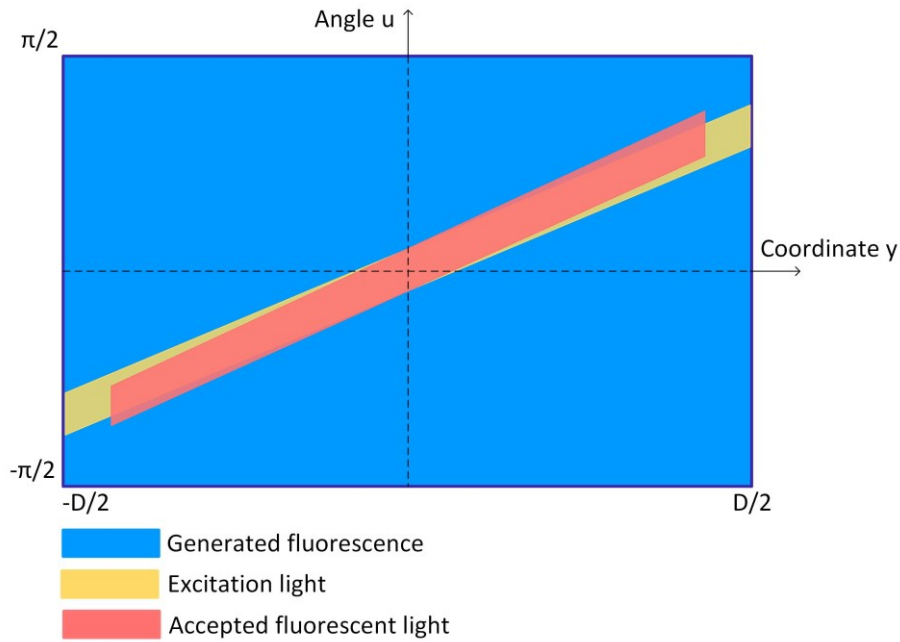


Fig. 4-2. Sketch of the spatial and angular overlap of the excitation (yellow), fluorescence (blue) and detected (red) light in the phase space in one dimension. The efficiency of transfer is given by the common area between the detection and fluorescent light, which is seen in red.

Due to the small acceptance angle, a small phase space overlap between the generation and acceptance of the fluorescence light is usually observed. The low geometrical coupling efficiency between the excited light cone and the acceptance light tube by one pixel is approximated by a simple model. As shown in Fig. 4-2, the coupling efficiency in the 1D case is $g_{1d} = \Delta u \cdot \Delta y$, while in the 2D case it is $g_{2d} = \Delta u \cdot \Delta v \cdot \Delta x \cdot \Delta y = \Delta \Omega \cdot \Delta S$. Besides the geometrical phase space coupling efficiency, the absolute power of autofluorescence light collected by the detector is also dependent on the physical autofluorescence coefficient of the glass material and the thickness of the z-slice, which together determine the probability for an excitation photon to be converted into a fluorescence photon in this z-slice. Finally, if only a single pixel is considered and the constant factors are collected in a normalization constant N , the transfer efficiency of autofluorescence light in the general 2D case is calculated as follows:

$$\frac{P}{P_0} = N \cdot \alpha \cdot \eta_{fluo} \cdot \sum_j^{n\text{-planes}} \left(\frac{NA_{excitation}^2}{NA_{full}^2} \right) \Delta z_j \Delta \Omega_j \frac{S_{overlap}}{S_{excitation}}. \quad (4.1)$$

Here $P_0 NA_{excitation}^2 / NA_{full}^2$ is the power of the light source that reaches the slice j , Δz_j is the thickness of the slice no. j , $\Delta \Omega_j$ is the solid angle of the pixel as seen from the center of slice j , $S_{overlap} / S_{excitation}$ is the ratio of the excitation area which overlaps with the acceptance area. η_{fluo} is the physical efficiency of autofluorescence light generation and α is the absorption coefficient. The last two terms describe the geometrical area overlap in the phase space as explained before. In this formula, the sum represents the total geometrical factor of the transfer efficiency of the lens, while the pre-factor involves the physical parameter of the material.

As for the efficiency of the simulation, it is found in practice that with a Matlab-based implementation for the photometric calculation and raytrace, the increase in speed for typical microscope lenses are in the range of 10^4 compared to conventional brute-force methods, which in this case is the volume scattering calculation of OpticStudio. This enables the optical designer to incorporate this kind of straylight analysis into a merit function for optimization. It gives the opportunity for fast identification of the microscope lenses, which shows a particularly high background intensity level due to autofluorescence of the glasses.

4.1.2 Simulation results and error analysis

If the phase space model is considered in more detail, a rigorous photometric analysis of the areas and angles is necessary, which should take the real aperture size of each element into account. The corresponding $NA_{excitation}$ and $S_{excitation}$ are obtained by tracing paraxial rays from the object plane to the edges of the elements, therefore it is limited by the lens apertures and stop size. $S_{overlap}$ is obtained by reverse aiming of paraxial rays from the pixel to the edges of the elements and then calculating the overlapping area with the excitation area. Here the real edges of the elements need to be considered rather than the pupils because rays that are not able to pass the pupil in the unscattered path can reach the detector due to the change in direction during the scatter event. Additionally, the calculated angles or areas with respect to each preceding element must be compared and the minimum value should be taken. Vignetting effects of skew rays are taken into account by aiming paraxial rays to the real edges of the lenses. As an example of this rigorous photometric analysis, Fig. 4-3 shows the layout of a plastic camera lens as well as the simulated geometrical autofluorescence contribution of each element.

The approximations used in this example are paraxial raytrace, circular approximation of vignetted pupils, a finite discretization of the lens volume assuming constant height and optical paths for all rays inside every slice, monochromatic excitation of the fluorescence light, uniform illumination and identical acceptance angle for all y -positions in each slice.

The determination of the number of z-slices depends on the thickness of the lenses and the largest ray angles inside the lens volume. If the shapes of the lenses are complicated with strongly curved surfaces or additional chamfer, the size of the considered slices is modified correspondingly. Furthermore, a depletion of the excitation light, coating effects and absorption are neglected, and all lenses are considered separately. In this section, we try to evaluate the influence of the major approximations on the accuracy of simulation.

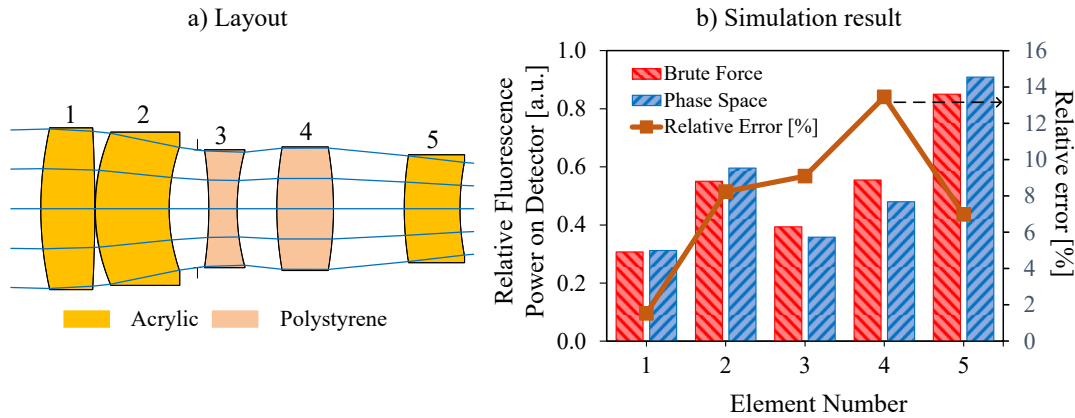


Fig. 4-3. Layout (a) and simulation result (b) of a plastic projection lens. The simulation results are calculated geometrically with the phase space and Monte Carlo raytracing without considering the autofluorescence coefficient of the plastic material.

The simulation results shown in Fig. 4-3(b) are calculated using the phase space model and the brute-force method of Zemax. From the orange curve we can conclude that the relative errors in most elements are within 10%, which is accurate enough for straylight analysis. However, the accuracy of the phase space model is strongly dependent on the structure of the optical system. As an example, Fig. 4-4 shows the simulation results of the autofluorescence effect of a high-NA microscope lens. From the orange curve in Fig. 4-4(b) we can conclude that the relative error of the phase space method is extremely large for the cover glass, immersion oil and the first element while it decreases rapidly in the middle and rear group, the fluctuation observed in the rear group is mainly statistical error due the limited number of rays that can be traced in the brute-force method. The large error observed in the front group originate from the large marginal ray angles in this lens group, which induces errors when rays are traced paraxially. In contrary, from the layout of the plastic lens in Fig. 4-3(a), we can see that the aperture angles of the ray bundle inside the plastic lens are much smaller compared to that in the front and rear groups of the microscope lens in Fig. 4-4. Therefore, the aperture angle is a critical factor which influences the accuracy of the phase space method. In this section, we try to evaluate the errors of the phase space method, and the possibilities to improve the accuracy of the method are investigated.

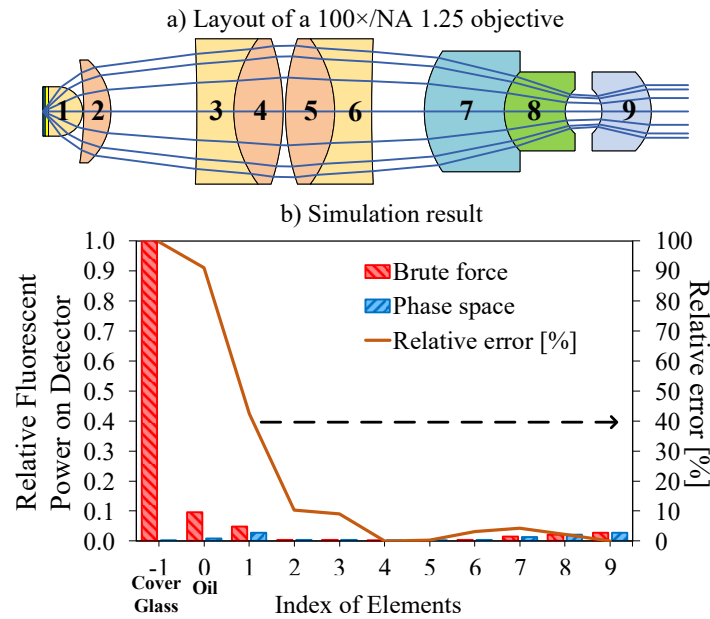


Fig. 4-4. Layout and simulation result of a 100 \times /NA 1.25 microscope lens [21]. The simulation results are calculated by the phase space and brute-force method without considering the autofluorescence coefficient of the glass, the orange curve corresponds to the relative error of the phase space method. Color-coding of the layout is not related to the autofluorescence coefficient of glass.

4.1.2.1 Paraxial approximation

In the phase-space-based model, the ABCD-based paraxial ray tracing is used instead of real ray tracing for faster calculation. However, large errors are expected for paraxial ray tracing of rays with large angles, which is typically seen in the front groups of microscope lenses. Here we try to evaluate the errors induced by paraxial ray tracing in the front group.

Fig. 4-5 shows the excitation area, acceptance area and acceptance angle calculated by paraxial and real ray tracing in the first element of the microscope lens in Fig. 4-4(a). From Fig. 4-5(a) we observe that the excitation area calculated by paraxial ray tracing and real ray tracing are identical, this is mainly due to that vignetting at the other elements has no impact on the excitation in the first element, and the object is assumed to be emitting light isotropically, the excitation area is only limited by the edge of the first element. Therefore, the height of the excitation area is simply the semidiameter of the first element at different z -locations.

Furthermore, from Fig. 4-5(c) it is important to note that the acceptance angles calculated by paraxial ray tracing and real ray tracing are in perfect agreement, despite the fact that here the acceptance angle is calculated for the full size of a $20 \times 20 \text{ mm}^2$ detector. Therefore, the accumulated error of reverse paraxial ray tracing from the detector plane to the first element is negligible for the calculation of the acceptance angle. From Fig. 4-5(c) we can also see that the acceptance angles are not extremely large, even for the smallest z location, where it is nearest to the image conjugate of the detector plane, the

half acceptance angle is only 9° for the full detector, which is not a large deviation from the paraxial range. In this figure it is obvious that the error increases with the acceptance angle. Therefore, we can infer that the deviation from the paraxial range determines the error of the paraxial based calculation.

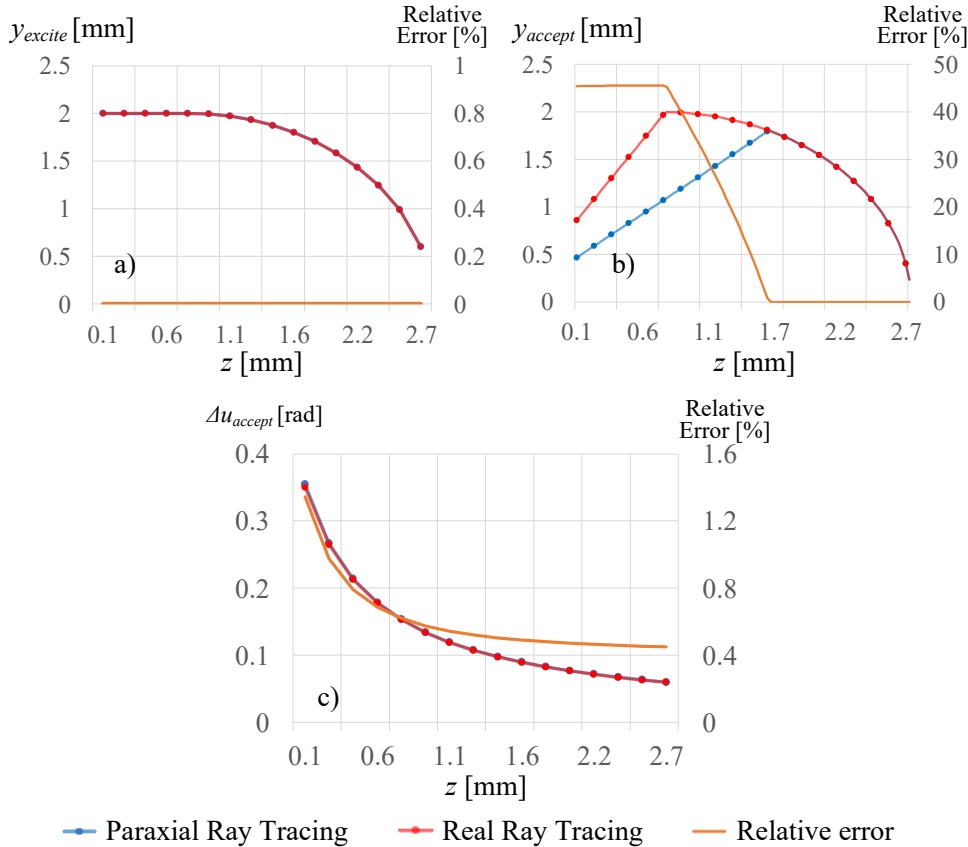


Fig. 4-5. Excitation area (a), acceptance area (b) and acceptance angle (c) calculated by paraxial ray tracing and real ray tracing for the first element.

Besides the excitation area and acceptance angles, which can be accurately calculated by paraxial ray tracing, the acceptance area is the worst offender. In Fig. 4-5 (b) we observe large deviation between the two curves, especially for the front part of the element. The largest deviation is around 40%, considering the actual acceptance area is the square of the height, the actual error of the acceptance area is even larger. This large error is due to the fact that the acceptance area in element 1 is not determined by its own edges, but the edge of element 6, which truncates the inverse ray bundle that is traced backwards from the image plane to the lens. In paraxial ray tracing, the marginal ray, that goes through the edge of element 6, is traced backwards from the image plane through the whole system to the first element, the accumulated error of paraxial raytracing and the large marginal ray angles in the first element result in large error of the marginal ray heights and angles in the first element.

Because the error originates from the assumptions of paraxial ray tracing, they are unavoidable if the paraxial approximation is poor. Therefore, the only solution to correct these errors is to incorporate real ray tracing in the calculation of the acceptance area for

critical elements. In the improved simulation code, an iterative ray aiming method is implemented to find the largest angle of the marginal ray which is not vignetted by succeeding elements. The acceptance areas in these elements are then calculated based on this marginal ray. As a consequence, the calculation time is slightly increased by this additional procedure, but considering that this subroutine is only called once for each critical element, the additional calculation time is acceptable (~ 0.3 s for each critical element).

4.1.2.2 Apodization

Besides the paraxial approximation, the most critical approximation is that of uniform illumination. Uniform illumination assumes that the illumination is uniform in every slice along the x - y direction, this is in general true for middle and rear groups of the microscope lenses, but invalid in the front group, which is near to the object. Due to the large aperture angles in the front group, the illumination is far from being uniform considering the projection with the cosine factor. The aberrations may also influence the uniformity but for the front group the cosine factor is dominant. Fig. 4-6 shows the intensity distribution on two slices in the first element of the above-mentioned $100\times/\text{NA } 1.25$ lens, we observe that the intensity at the center of the slice is one order of magnitude higher compared to that at the edge. Therefore, the uniform illumination approximation is the main origin of the large error of the first element as we observe in Fig. 4-4(b). In order to correct for this error, the real intensity distribution should be considered in the critical elements. In the improved simulation code, the real intensity distributions in critical elements are calculated by launching rays on an equidistant mesh and then calculating the intensity distribution in each slice from the distorted mesh. Then the excitation area is weighted by the intensity distribution during calculation of its overlap with the acceptance area.

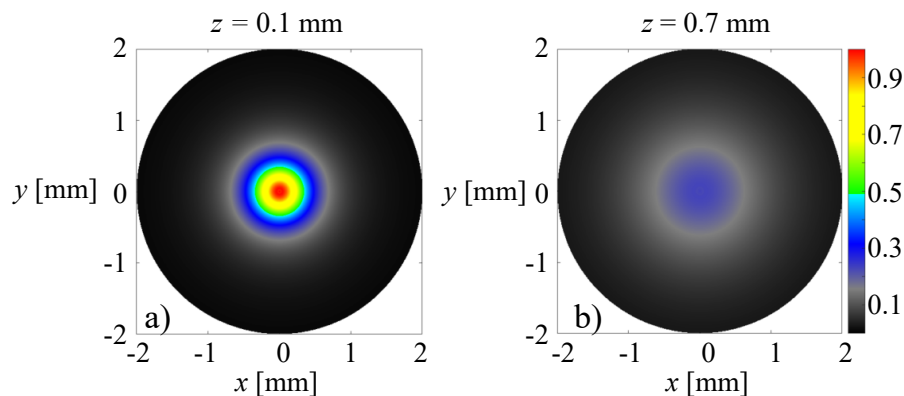


Fig. 4-6. Intensity distribution in the first element of the $10\times/\text{NA } 1.25$ microscope lens at $z = 0.1$ mm (a) and $z = 0.7$ mm (b).

Another consequence of the large aperture angles in the front group is that the assumption of equal optical paths of all rays within a slice is no longer valid. The rays at the outer part of the element have larger angles and therefore go through longer optical paths within the slice, which means they are more likely to be scattered compared to the rays at the

center. However, considering that in the front group the energy is concentrated at the center, the extended optical path for outer rays will only have a minor influence on the final autofluorescence contribution. Therefore, if the excitation area is weighted by the real intensity distribution, the approximation of equal optical paths will not induce large errors in the simulation result.

4.1.2.3 Vignetting and aberration

In the phase space model, the vignetted pupils are assumed to be circular. This is in principle not true for truncated circular pupils, however, since microscope lenses are aperture dominated systems, the deviation of the vignetted pupil from a circle is not large.

Additionally, it is also assumed that the acceptance angles are constant at all y -positions in a slice. However, the acceptance angle at each position of the slice is the angular opening of the intermediate image of the detector seen from this position. Therefore, the angular opening varies with the y -position due to the projection effect. Additionally, the intermediate image of the detector is blurred by the aberrations of the system after this slice, and the edge of the intermediate image also suffers from vignetting. Therefore, the angular opening for each y position is influenced by the transverse aberration and vignetting of the system at the corresponding pupil coordinate. In order to visualize the deformation of the acceptance domain in the front group, we have to choose a system without perfect index match in the object, so that the aberration from the rear part of the system is large to correct for the aberration induced by index mismatch. Therefore, we substitute the glass used in the first element of the lens shown in Fig. 4-4(a) by the glass N-LASF41, which is a flint glass with a refractive index of 1.84, then the system is re-optimized for equivalent performance as the original design.

The blue area in Fig. 4-7 shows the ideal acceptance domain of a slice at $z = 1.6$ mm in the first element of the above-mentioned microscope lens without considering the aberrations and vignetting.

First of all, we can see from the above figure that its shape is not a perfect parallelogram but bends at the right and left edges, where the width of the shape is also narrower. This is due to the projection effect and the shape of the curve being an arc tangent function. The red area shows the real acceptance domain with all aberrations and vignetting considered, and the impact of aberrations and vignetting on the acceptance domain can be obtained by comparing the blue and red area. Due to spherical aberration, the shape is slightly twisted clockwise, while strongly narrowed at the right and left boundary of the area due to vignetting and coma. Although aberrations influence the shape of the phase space acceptance area, it is obvious that the size of the area is not tremendously changed for this case where large spherical aberrations and coma are present. Therefore, it is reasonable to neglect the aberrations and vignetting when calculating the acceptance domain in phase space. Furthermore, the upper and lower boundaries of the ideal

acceptance area can be analytically expressed by arc tangent functions, thus the exact size of the acceptance area can be directly calculated by integration. However, considering that in the first element most energy is concentrated at the center, exact calculation of the acceptance angle at the edge of the element is not expected to yield a large improvement in accuracy and is therefore not yet implemented in the simulation code.

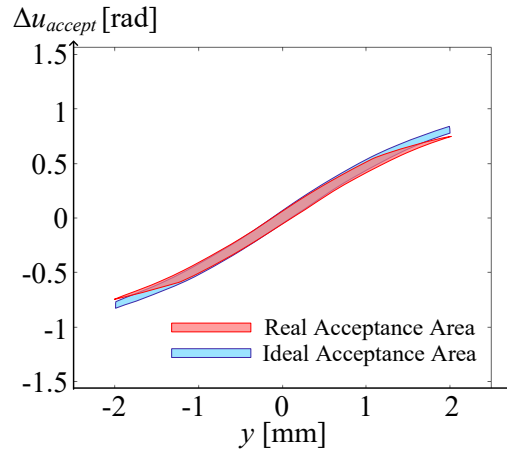


Fig. 4-7. Real phase space acceptance area and ideal acceptance area without aberration and vignetting.

4.1.2.4 Discretization of detector

Another important aspect to be considered is the discretization of the detector. Fig. 4-8 shows the dependence of the fluorescence intensity on the image height.

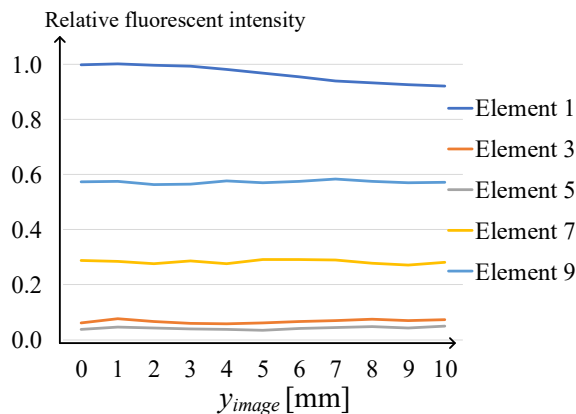


Fig. 4-8. Dependence of the fluorescent intensity on image height.

It can be inferred from the above figure that the dependence is extremely weak. Even for the first element where it is near to the conjugation position of the detector, no obvious dependence of the fluorescent intensity on the image height can be observed. This is because the large NA of the microscope lenses results in extremely small depth of field, even though the fluorescent centers in the first element are near to the conjugation position, they cannot be imaged on the detector. Therefore, the impact of autofluorescence effect on the image is a nearly uniform intensity offset and thus fine discretization of the detector is not necessary to improve the accuracy. Considering the large computational effort required by discretization of the detector, we can ignore the discretization of the detector without loss of accuracy. Based on the above discussion, that the most critical elements,

considering the accuracy of the phase space model, are the ones in which marginal ray angles are large, which corresponds to the error curve in Fig. 4-4(b). The main origins of the errors are the non-uniform intensity distribution as well as the spatial acceptance area.

4.1.2.5 Improved method

According to the discussion in the last section, the errors of the paraxial phase space method can be eliminated by the implementation of real ray tracing in critical subroutines and weighting of the excitation area for phase space coupling. In order to achieve this goal, the phase space method is improved by applying real ray aiming and apodization in the critical lens groups. To evaluate the properties and in particular to check the accuracy of the improved method, test calculations are performed with OpticStudio and the improved phase-space-based simulation code. The corresponding simulation result for the lens in Fig. 4-4(a) is shown in Fig. 4-9, in which it is seen, that both approaches are only slightly different.

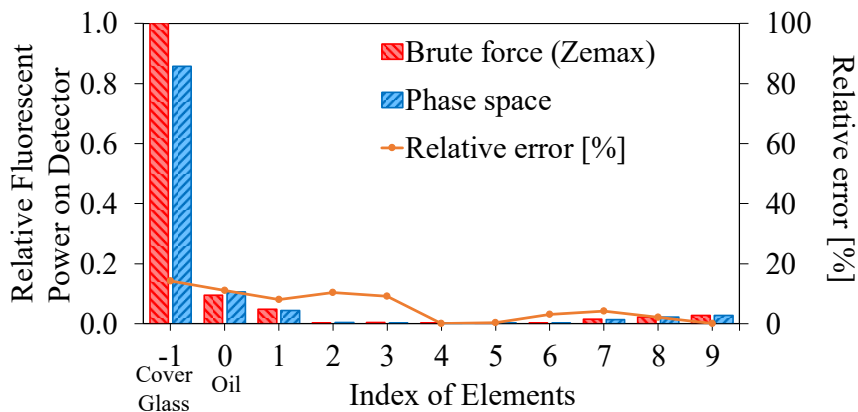


Fig. 4-9. Comparison of phase-space-based model with Zemax for the example system in Fig. 4-4(a). The calculation with the paraxial model was performed for full NA.

Due to the implementation of real ray tracing in the calculation of the acceptance area and the intensity distribution, a large difference is no longer seen in the first group. Although the error of the front group is still larger than the middle and rear groups, the RMS value of the overall relative error is within 10%, which is negligible considering the statistical error of the brute-force method and the uncertainty of the material parameters. The number of rays in the brute-force simulations is selected to get a statistical uncertainty of smaller than 10%. With these numbers Zemax needs 10 hours for the computation, while the phase-space-based tool only needs 10 seconds (Intel E5-2690 V4 $\times 2$ @2.6GHz).

4.2 Application and practical consequences in optical design

4.2.1 Overview

According to Eq. (4.1) and the quasi-analytical model introduced in Sec. 3.2, the autofluorescence contribution from an element of a microscope lens is determined by its thickness, internal marginal ray height and material data, while it is scaled by the system etendue. To investigate the autofluorescence effect of different types of microscope lenses, we have applied the phase space method to simulate the autofluorescence effect of a variety of microscope lenses with different NA and magnification from different vendors. In this section, we summarize all simulation results and evaluate the dependence of the autofluorescence intensity of a microscope lens on the etendue, working distance and structure of lens groups. With the strategic analysis of the simulation result, we can determine the critical parameters and lens group structures that have large impact on the autofluorescence effect, based on which the corresponding optimization strategies for optical design can be concluded.

Fig. 4-10 shows the relationship between the geometrical autofluorescence intensity and the system etendue for a selection of lenses. Here the etendue of a microscope lens is defined as $y^2 \cdot NA^2$, where y is the object height and NA is the object space numerical aperture. The etendue calculated in this way differs from the standard definition of etendue by a factor of π^2 . In this paper, we use this definition instead of the standard definition because it is easier to be calculated so that the lens designers and users can more conveniently estimate the autofluorescence intensity of their microscope lenses based on the results shown in the figures. The unit of the etendue is defined as $[\text{rad}^2 \cdot \text{mm}^2]$ rather than $[\text{m}^2 \cdot \text{sr}]$ to distinguish from the standard definition of etendue.

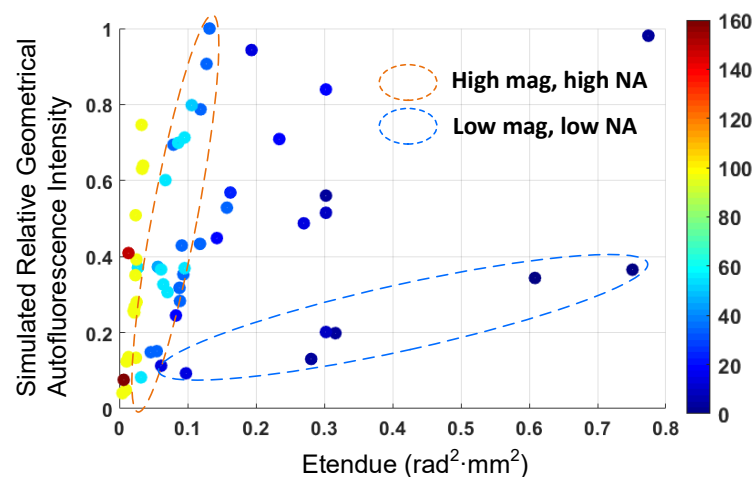


Fig. 4-10. Simulated relative geometrical autofluorescence intensity vs. system etendue, the maximum value has been normalized to 1, the magnifications of the microscope lenses are indicated by the color code. The etendue is defined as $y^2 \cdot NA^2$, where y is the object height and NA is the object space numerical aperture.

In Fig. 4-10, 60 microscope lenses are selected from a large lens database [22] in such a way that most commonly seen microscope lens designs from the German and Japanese vendors are covered, a detailed list of the microscope lenses in the database can be found in the appendix of [22]. As seen from Fig. 4-10 it is obvious that the points are rather scattered, which means that a clear dependence of the autofluorescence intensity on the system etendue is not seen. However, if we consider the color code, which indicates the magnification of the microscope lenses, we find that the sensitivity of the autofluorescence effect to the system etendue is strongly dependent on the magnification, which normally has a positive correlation with the NA and complexity of microscope lenses. In Fig. 4-10, the two dashed ellipsoids encircle microscope lenses with similar NA and magnification, we find that if we categorize the microscope lenses according to the NA and magnifications, the linear dependence of autofluorescence intensity on the system etendue is much more pronounced. Fig. 4-11 shows the dependence of autofluorescence intensity on etendue for high NA, high magnification lens and low NA, low magnification lenses in separated representations. In Fig. 4-11 we observe that in each lens category, the autofluorescence intensity is roughly linearly dependent on the system etendue as indicated by the best fit lines, while the slopes of the two lines are very different from each other as a result of different sensitivity to the system etendue.

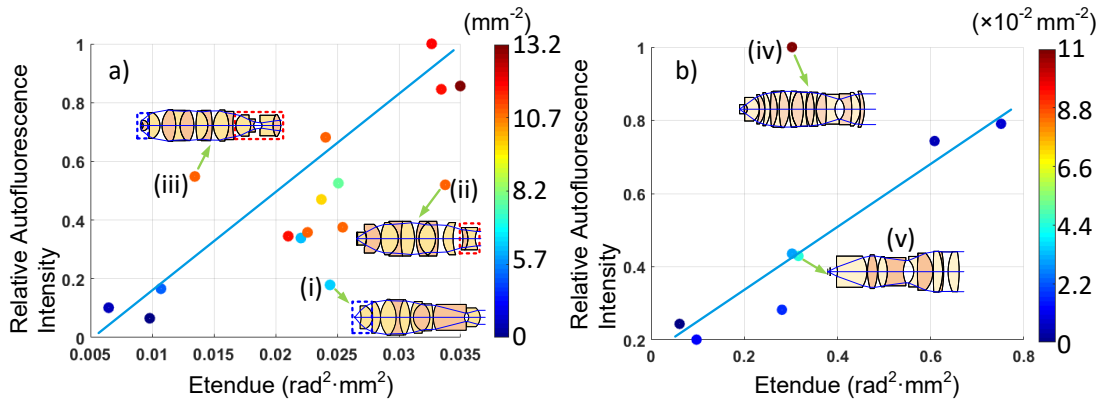


Fig. 4-11. Geometrical autofluorescence intensity vs. system etendue for high NA, high magnification (a) and low NA, low magnification (b) lenses. The color code represents $1/w_d^2$, where w_d is the working distance of the microscope lens. The etendue is defined as $y^2 \cdot NA^2$, where y is the object height and NA is the object space numerical aperture.

Although most of the lenses in Fig. 4-11 are scattered closely around the best fit lines, we also notice several exceptions. In Fig. 4-11(a) we note that there are three lenses (i,ii,iii) located far from the best fit line, among which lens (iii) [23] is far above this line while lens (i) and (ii) [24, 25] are far below this line. The corresponding layout of these three lenses are shown in Fig. 4-11(a) with the working distance indicated by the color code. The blue and red dashed rectangles indicate the critical front and rear groups which dominate the autofluorescence intensity of each lens. From the layouts in Fig. 4-11(a), we can see that the lens (iii) has a much shorter working distance compared to lens (i), and a

much more complicated rear group compared to lens (ii). The rear group of lens (ii) is only composed of a thin cemented element, while that of the lens (iii) is the Gauss-type rear group composed of thick meniscus lenses, which will be shown in Sec. 4.2.3 to be a very sensitive structure in terms of the autofluorescence effect. Similarly, among the low NA, low magnification lenses we also observe a lens (iv) [26] far from the best fit line, as shown in Fig. 4-11(b). Compared to lens (v) [27] which is located closely to the best fit line, lens (iv) has a shorter working distance and the critical Gauss-type rear group. Therefore, for microscope lenses with similar NA, magnification and working distance, the autofluorescence intensity of a microscope lens is roughly linearly dependent on the system etendue, while the slope of the linear dependence is determined by the structures of the critical lens groups. A more detailed analysis of the critical lens groups will be presented in the next sections.

4.2.2 Autofluorescence coefficients of glass

From Eq. (4.1) we learn that the autofluorescence coefficients of glass directly scale the geometrical autofluorescence contribution from each element, therefore it has a great impact on the real autofluorescence intensity of microscope lenses. The autofluorescence coefficients greatly vary for different types of glasses, which is mainly due to their different chemical compositions. Therefore, the calculation of real autofluorescence intensity should always incorporate the real autofluorescence coefficients of the glasses used in each element. However, unlike the refractive indices and dispersion, the autofluorescence coefficient of optical glass is not a standard property which is always provided by the glass vendors, and the amount of the available information is rather limited. As a consequence, a comprehensive analysis based on the real autofluorescence coefficients of all optical glasses is not possible, and here a qualitative analysis based on the limited information provided by the glass manufacturer Schott [10, 28, 29] is presented.

So far, we have collected the autofluorescence coefficients of 48 types of Schott glasses with various refractive indices and Abbe numbers. Fig. 4-12(a) shows the positions of these glasses in the glass map and the color code corresponds to their relative autofluorescence coefficients with an excitation wavelength of 365nm. According to Fig. 4-12(a), it is obvious that the glass with the highest autofluorescence coefficients is all lead-containing glass ('non-N-type'), while the autofluorescence coefficients of corresponding lead-free glass ('N-type') are at least one order of magnitude lower. Therefore, we can conclude that the lead ions play a very important role in the generation of fluorescent light for excitation wavelength of 365nm (i-line), and the use of lead-free glass is much more preferred in the microscope lenses designed for fluorescence microscopy. Correspondingly, Fig. 4-12(b) shows the autofluorescence coefficients of the lead-free glasses. With the absence of lead, the autofluorescence coefficients are

determined by the glass matrix and the content of other substances such as the rare earth elements. From the figure it is obvious that the autofluorescence coefficients of glasses with high refractive indices and dispersion are much larger than that of low refractive, low dispersion glasses, while the behavior of the glasses in between these two regions is not clear. By comparing Fig. 4-12 (a) and (b), we notice that some special glasses, such as LASF35, do not have a lead-free version, while this glass is necessary in some lenses due to its large refractive index. Therefore, the use of lead-containing glass cannot be completely avoided in microscope lenses.

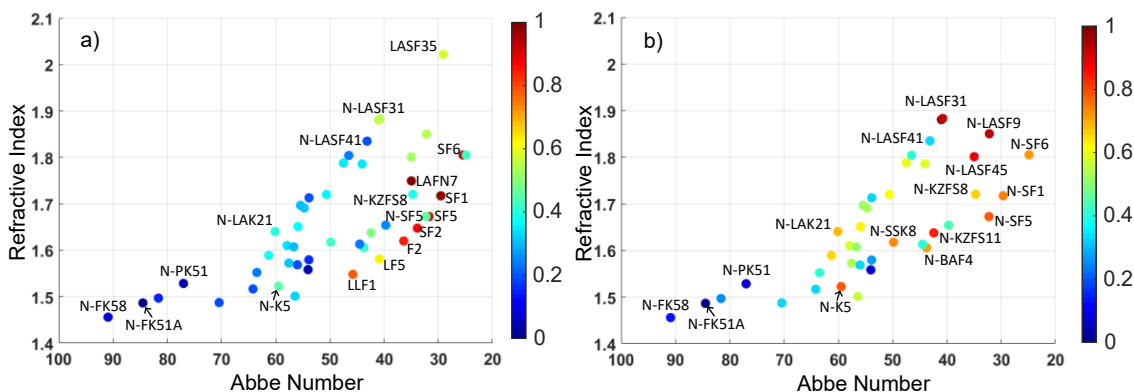


Fig. 4-12. Autofluorescence coefficients of optical glasses with an excitation wavelength of 365nm. a) shows all glasses with available data while b) shows the lead-free ('N-type') glass, the color code corresponds to the logarithmically-scaled autofluorescence coefficients, which are normalized by the maximum value in each plot. The number of glass types in b) is less than that in a) since the glasses shown in b) is a subset of that in a).

Besides the refractive index and Abbe number, another important property of optical glasses is the partial dispersion, which is a key factor in the design of microscope lenses to correct the secondary spectrum. Fig. 4-13 shows the correlation between the anomalous partial dispersion ($\Delta P_{g,F}$) and autofluorescence coefficients.

From Fig. 4-13, we see that there is no clear dependence of the autofluorescence coefficients on the anomalous partial dispersion. The fluoro-phosphate crown glass (N-PK, N-FK series) shows positive anomalous partial dispersion and low autofluorescence coefficients, while the short flint glass (N-KZFS series) shows negative anomalous partial dispersion and low autofluorescence coefficients. The high refractive index lanthanum dense flint glass (LASF series) shows large autofluorescence coefficients.

Due to the different functionalities of the three lens groups of microscope lenses, there usually exist certain preferences of refractive indices and dispersion for each lens group. Considering that the autofluorescence coefficients are related to the refractive indices and Abbe numbers, it is necessary to consider the different glass types which are used in each lens group. Fig. 4-14 shows the relative frequency of used glasses in the glass map for the three lens groups.

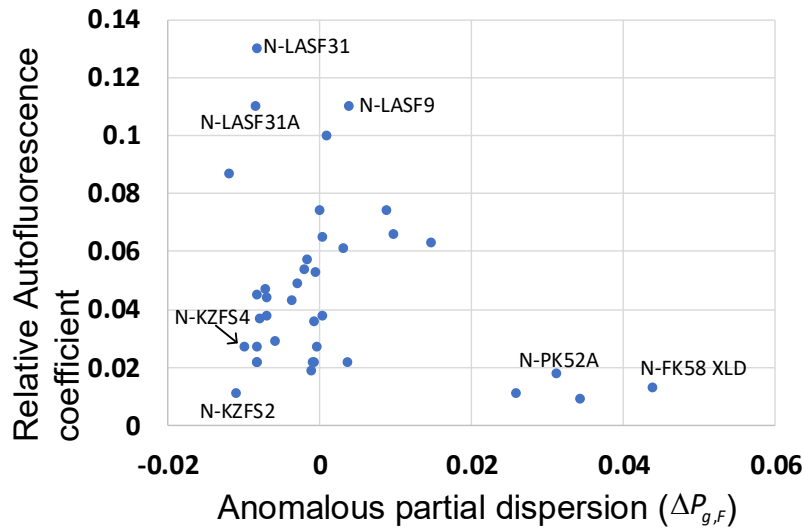


Fig. 4-13. Dependence of autofluorescence coefficients on the anomalous partial dispersion ($\Delta P_{g,F}$) for lead-free ('N-type') glasses.

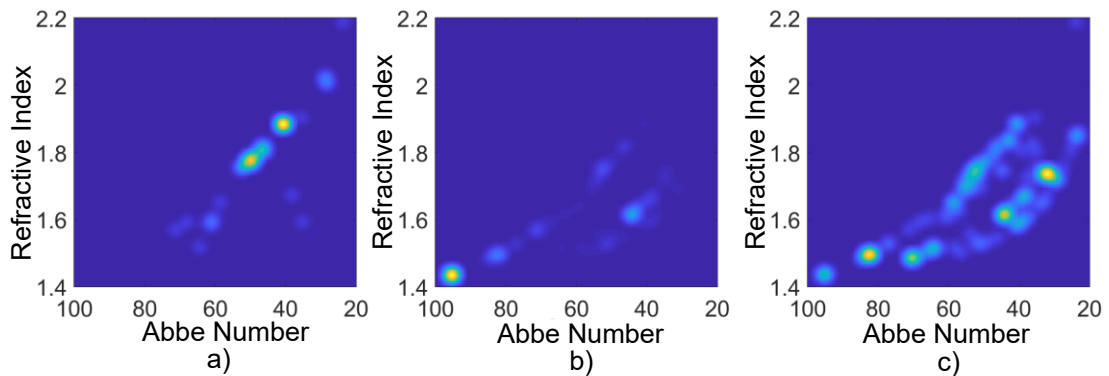


Fig. 4-14. Relative frequency of use of glasses in the glass map for the front (a), middle (b) and rear (c) groups. Gaussian smoothing is applied to each figure and the peak values are normalized to 1.

From Fig. 4-14 we can see that in the front group, glasses with high refractive indices and dispersion are mostly used, the refractive indices of the used glass are concentrated between 1.7 and 2.0, which is necessary in high NA lenses to quickly reduce the NA in the front group. Meanwhile, in the middle group, most of the glasses has low refractive indices and dispersion. In the rear group, the glass distribution is more scattered, high refractive index glasses are used mostly in the thick meniscus lenses for field flattening and spherical aberration correction, while low refractive index fluoro-phosphate crown glass with large anomalous dispersion is used for the correction of the secondary spectrum. Therefore, from a material point of view, the autofluorescence contribution from the middle group is less critical because glasses with low refractive indices and dispersion usually have low autofluorescence coefficients, while the autofluorescence contributions from the front and rear groups are larger due to the use of high refractive index glass.

4.2.3 Critical lens groups

In the previous section, we have found that the overall autofluorescence intensity of microscope lenses with similar NA and magnification is approximately linearly dependent on the system etendue, but the contribution from some specific lens groups, which are usually the front and rear groups, are usually large and has a great impact on the autofluorescence intensity of individual lenses. Here we define these lens groups, which contribute to a significant fraction of total autofluorescence, as critical lens groups. In this section, we analyze the autofluorescence effect in critical lens groups both geometrically and physically with the aim of finding the origins of the autofluorescence contribution from the critical lens groups, and thus provide a theoretical basis for the methods to reduce the autofluorescence effect, which will be discussed later.

In order to investigate the autofluorescence contribution from different types of lens groups, here we take lens (i) and lens (ii) from Fig. 4-11(a) as examples because these two lenses have similar magnification, NA and middle group, but different front and rear groups (NAs of lens (i) and (ii) are 1.0 and 1.1, magnifications of both lenses are 63). Fig. 4-15(a,b) again show the layouts of these two lenses, but with the autofluorescence coefficients of the glass indicated by the color code.

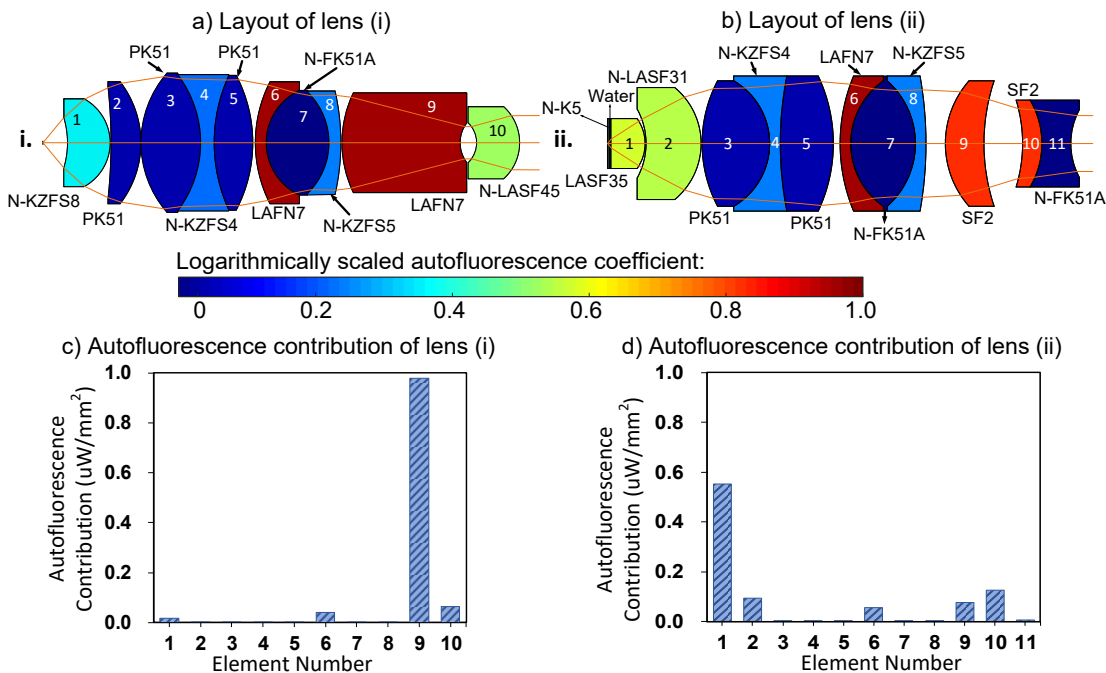


Fig. 4-15. Layouts and autofluorescence contribution from the elements of lens (i) and (ii). The light source is assumed to be an on-axis point source located at the object plane, emitting excitation light at 365nm isotropically in all directions. The power of the light source is 1 W. The fluorescent light is measured on the detector plane, the size of the detector is 11mm \times 11mm, the autofluorescence contribution from an element is characterized as the irradiance of fluorescent light on the detector. The color coding shows the logarithmically scaled relative autofluorescence coefficients, and the maximum value is normalized to 1. The results are based on a tube lens with a focal length of 200mm.

Fig. 4-15(c,d) show the corresponding autofluorescence contribution from each lens element, for which the absolute autofluorescence coefficients of glass are considered. By comparing these two figures, we find that the autofluorescence contribution from the middle groups, which are elements 3-8 in both lenses, is similar with each other as a result of similar etendue and structure, while large differences between the two lenses can be observed in the front and rear groups, which results from the different structure and material selection of these lens groups.

4.2.3.1 Front group

In Fig. 4-15, we observe that the geometrical autofluorescence contribution from the front group of lens (i) is much smaller than that of lens (ii), and the main difference between these two lens groups is the working distance. Therefore, the autofluorescence contribution from the front group should be largely dependent on the working distance. If the NA of a microscope lens is fixed, the working distance and the focal power of the front surface directly determine the marginal ray heights inside the first element, which, according to the analytical model, has a large impact on the geometrical autofluorescence contribution. According to Eq. (3.6), it has been derived that the acceptance solid angle of the generated fluorescence in an z -slice is determined by the refractive index and marginal ray height inside the z -slice. Considering that the focal power of the front surface of microscope lenses is usually very small, the marginal ray height in the first element is mainly determined by the working distance as $h = d \cdot \tan u$, where d is the working distance and u is the marginal ray angle. Substituting the above equation into Eq. (3.6), we get

$$d\Omega_o = \left[\pi p_i^2 (n_i u_i)^2 \frac{1}{\tan^2 u} \right] \cdot \frac{1}{n^2 d^2}. \quad (4.2)$$

From the above equation, we learn that the autofluorescence contribution from an element is inversely proportional to $n^2 d^2$. The analytical result is validated against simulation results in Fig. 4-16 by showing the dependence of the simulated geometrical autofluorescence contribution from the first elements on the working distance and refractive index. In the figure we find that a linear dependence of the geometrical autofluorescence contribution from the first element on $1/n^2 d^2$ can be observed, which agrees with the analytical derivation.

Considering that the thickness of the front group is usually much larger than the working distance, the marginal ray heights in the front group greatly varies with the z -location. Therefore, rather than considering the front group as a whole, the z -dependence of the autofluorescence contribution from different part of the front group should be considered. This should be the case for any thick element that locates closely to the intermediate image, as it will be shown later in the discussion about the rear group. In the phase-space-based model, the elements of a microscope lens are divided into several thin slices along the z -direction and the autofluorescence contribution from each z -slice is calculated

individually, which enables us to directly analyze the z -dependence of the autofluorescence contribution from different parts of an element.

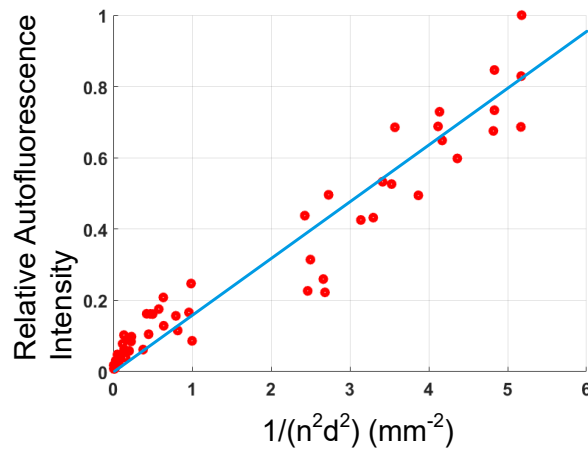


Fig. 4-16. Autofluorescence contribution from the first element vs. the index-weighted working distance.

Fig. 4-17(a) shows the z -sectioning of the first element of lens (ii). The red curve of Fig. 4-17(b) shows the analytically calculated autofluorescence contribution from each slice according to Eq. (4.2), which agrees with the simulation result of the phase space model, as indicated by the blue dots. The horizontal axis of Fig. 4-17(b) is the z -location of a slice relative to the intermediate image plane, which is indicated by the red dashed line in Fig. 4-17(a). From this figure, we can observe that the autofluorescence contribution from a z -slice decreases rapidly with the distance to the intermediate image, indicating that most of the autofluorescence contribution of the first element comes from a small area near the object. This can also be seen in the green curve, which shows the accumulated autofluorescence contribution. From the green curve we see that the accumulated autofluorescence contribution reaches 60% at $z = 1.05\text{mm}$, which means that 20% of the first element contribute to 60% of its autofluorescence contribution. Besides the first element, the 2nd element of lens (ii) also shows a large autofluorescence contribution, which results mainly from its large thickness, small distance to the intermediate image and the use of dense flint glass.

Considering the large geometrical sensitivity of the front group, glass with extremely low autofluorescence coefficient is necessary. However, from the perspective of optical design, in order to quickly reduce the NA, dense flint glass with large refractive index is necessary [30], but these types of glass usually have large autofluorescence coefficients. Therefore, it seems that both sides must be compromised in order to reach a balance, however, as it will be shown later, the dilemma can be overcome by lens splitting and judicious glass selection.

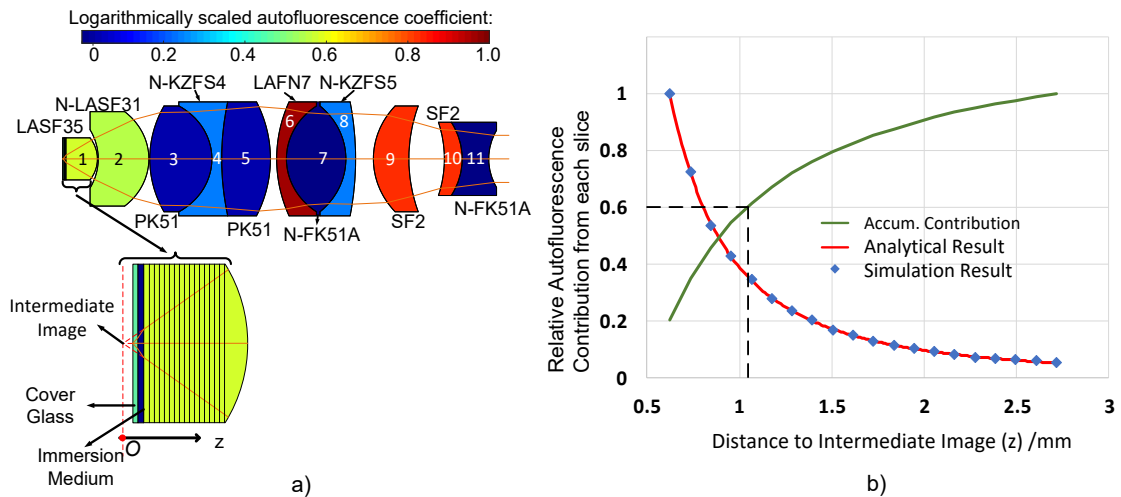


Fig. 4-17. Layout of lens (ii) and the z -sectioning of its first element in the phase-space-based simulation. The intermediate image in the first element is marked in order to indicate its distance to each slice. b) z -dependence of autofluorescence contribution from the first element. The horizontal axis is the z -location of a slice relative to the intermediate image plane, the red curve corresponds to the relative autofluorescence contribution from each slice, while the green curve corresponds to the accumulated autofluorescence contribution of all slices whose distance to the intermediate image is smaller than z .

Furthermore, another aspect that needs to be considered in the front group is the cover glass, whose autofluorescence contribution is usually neglected due to its small thickness. However, considering the very small distance between the cover glass and the object, its autofluorescence contribution can be dominant compared to the autofluorescence generated in the microscope lens. Fig. 4-18 shows the comparison between the autofluorescence contribution from different types of cover glass and first element of lens (ii). We can see that the autofluorescence contribution from the cover glass made of N-K5, for which lens (ii) is originally corrected, is 6 times larger than that from the first element, which is too high to be neglected. Meanwhile, we can also see that the autofluorescence contribution can be significantly reduced if a cover glass made of the glass N-FK51A is used, but the reduced value is still within the same order of magnitude compared to that of the first element. Nowadays borosilicate glass is more commonly used as the cover glass, although the glass manufacturers claim that this type of cover glass is ‘lowly fluorescent’ [31], exact values of the autofluorescence coefficients of these types of cover glass are unavailable. However, according to the datasheet provided by the glass manufacturer Schott [32], the autofluorescence coefficient of borosilicate glass is comparable to soda-lime glass in the UV region, meaning that the autofluorescence contribution from the borosilicate cover glass cannot be neglected.

The same problem also occurs for the immersion oil, which is also very close to the object. However, it is unfortunate that the autofluorescence property of immersion oil is even more difficult to access than glass and thus a more detailed evaluation of its impact on the autofluorescence intensity is not possible.

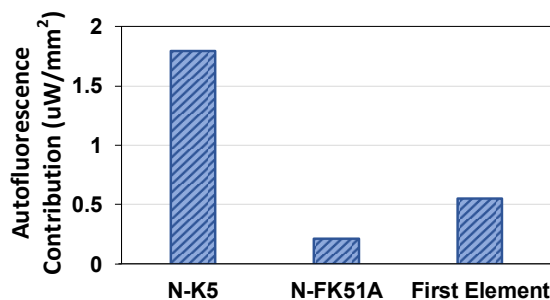


Fig. 4-18. Autofluorescence contribution from different types of cover glass in comparison with the first element of lens (ii), the thickness of the cover glass is 0.17mm.

4.2.3.2 Rear Group

It can be observed in Fig. 4-15 that the autofluorescence contribution from the rear group of lens (i) is much larger than that of lens (ii). The rear group of lens (i) is composed element 9 and 10, which are two thick meniscus lenses made of dense flint glass. The large autofluorescence contribution from this rear group is partially due to its large thickness and high autofluorescence coefficients of dense flint glass, while its focal power distribution is another key factor that leads to its large autofluorescence contribution.

The rear group of lens (i) is the so-called Gauss-type [22]. It features a symmetrical structure with its focal power distributed as positive-negative-negative-positive (+-+) among its four surfaces. This special power distribution results in small marginal ray height near the two inner surfaces, which, according to Eq. (4.2), leads to large autofluorescence contribution from this part of the lens. Similar to Fig. 4-17, Fig. 4-19 shows the autofluorescence contribution from different z-slices of element 9 in lens (i). From Fig. 4-19(b) we observe that the autofluorescence contribution from the z-slices of the rear group increases significantly as the distance to the intermediate image plane decreases, indicating that most of the autofluorescence contribution from element 9 is confined in a small area near the inner surface of the rear group. As it can be seen from the green curve, which shows the accumulated autofluorescence contribution, 35% of element 9 contributes to 60% of the autofluorescence contribution. The red curve corresponds to the analytical result calculated by Eq. (4.2). The agreement between the analytical and simulation results again validates the analytical model.

Based on the above discussion, we find that the Gauss-type rear group is not ideal considering the autofluorescence effect due to its large thickness and small marginal ray height near the inner surfaces. However, this kind of rear group is essential in high NA microscope lenses for color correction and field flattening. Additionally, the aberration correction requires the use of dense flint glass in the rear group, which, according to Sec. 4.2.2, further increases the autofluorescence contribution. Therefore, the optical design and the material selection of the Gauss-type rear groups should be done carefully to suppress the autofluorescence contribution, the details will be discussed later.

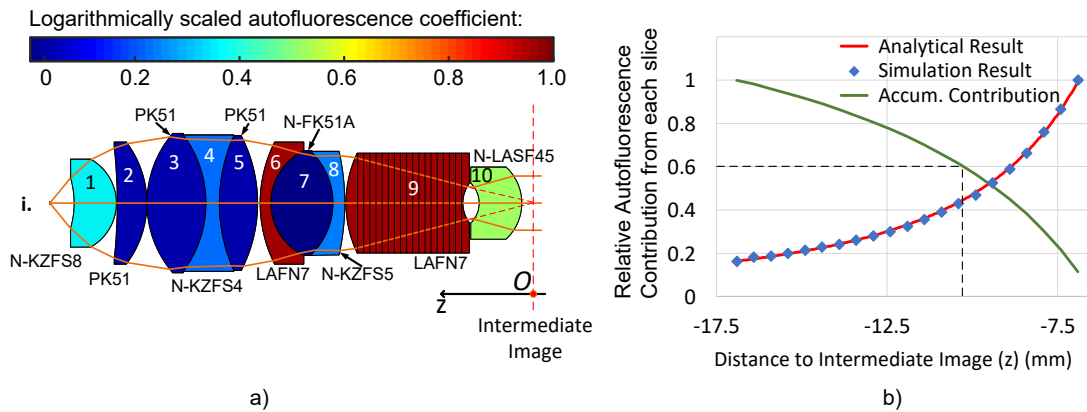


Fig. 4-19. a) Layout of lens (i) and z-sectioning of its rear element in the phase-space-based simulation. The intermediate image in the 9th element is marked in order to indicate its distance to each slice. b) z-dependence of autofluorescence contribution from element 9, the horizontal axis is the z-location of a slice relative to the intermediate image plane, the red curve corresponds to the relative autofluorescence contribution from each slice, while the green curve corresponds to the accumulated autofluorescence contribution of all slices whose z-coordinate is larger than z, with the intermediate image plane being the origin.

4.2.4 Suppression of the autofluorescence effect in microscope lens design

In the last sections we have discussed the influence of different system parameters on the autofluorescence effect and the autofluorescence contribution from different lens groups. Based on the above discussions, here we again take the two microscope lenses in Fig. 4-15 as examples and show how to effectively reduce the autofluorescence intensities of existing lens designs. The optimization of the microscope lenses as described below is within the spectral range from C-line (656nm) to F-line (486nm).

4.2.4.1 Glass substitution

According to the previous discussion about the cover glass, substituting the N-K5 cover glass by the glass N-FK51A can reduce the autofluorescence contribution from the cover glass by nearly 90%. Therefore, the most direct way to reduce the autofluorescence intensity of a microscope lens is to substitute the glass in critical lens groups by another type of glass with smaller autofluorescence coefficient. In Fig. 4-15 we see that the glass LAFN7, which has a very large autofluorescence coefficients, is used in both lenses and results in large autofluorescence contribution from the corresponding elements. The same can be seen in element 9 and 10 of lens (ii), in which the glass SF2 is used.

Besides these lens elements with highly fluorescent glass, special care also has to be taken for the critical lens groups, which, according to the discussion in the last section, has a lower tolerance for autofluorescence coefficient of glass due to their high geometrical sensitivity. For example, in lens (i), the autofluorescence contribution from element 10 is still large while the ‘N-type’ glass is already used in this element, the same can be found

in the 2nd element of lens (ii). Therefore, glass substitution should be performed for elements with highly fluorescent glass or high geometrical sensitivity.

Considering the choice of the substitution glass, the best substitution is usually found in the neighbors of the original glass on the glass map, ideally the nearest neighbor with lower autofluorescence coefficient should be used as the substitution. Fig. 4-20 shows the selection of the substitution glass. In Fig. 4-20 it is seen that proper substitution can be found for most of the glass, except for the glass LASF35 in the first element of lens (ii). In this element, a ‘N-type’ glass with extremely low autofluorescence coefficient is necessary to compensate for its high geometrical sensitivity. However, due its extremely large refractive index ($n = 2.02$, $v_d = 25.43$), it is impossible to find an adjacent ‘N-type’ glass in the glass map as a substitution. The refractive index of the nearest ‘N-type’ glass is too low to preserve the system performance without significant structural change. In this case a simple glass substitution for this element cannot reduce its autofluorescence contribution.

After the glass in the critical elements is substituted by low-fluorescent glass, the radii of curvatures of the lenses have to be re-adjusted to obtain similar performance as the initial system. The difference between the original and substitution glass determines the amount of effort that needs be taken to reoptimize the system, e.g., large difference in refractive indices leads to difficulty in suppressing the additional spherical aberrations while large difference in the Abbe number requires more effort on color correction. Considering the optimization strategy, a step-by-step optimization after each glass substitution is more preferred than a single optimization after all glass substitutions.

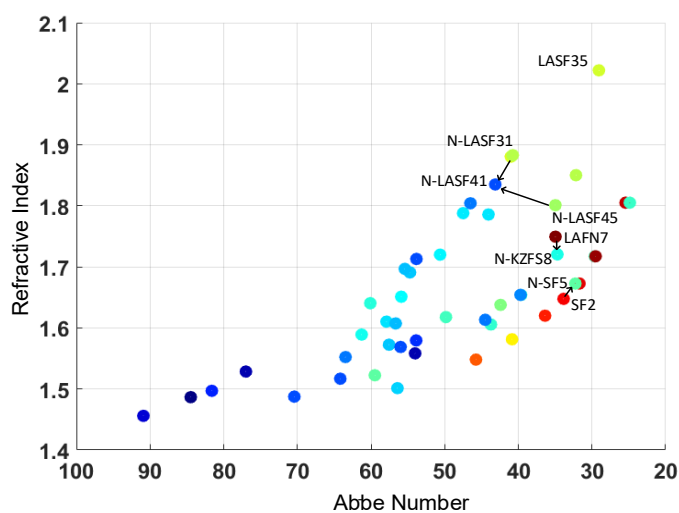


Fig. 4-20. Glass substitution based on the glass map.

Due to the relatively small difference between the original and substitution glass, we are able to obtain similar performance as the initial systems for both example lenses by re-optimization of the radii of curvatures after glass substitution. The corresponding layouts and autofluorescence contribution are shown in the Fig. 4-21, from which we see that the

structures of the optical systems remain almost unchanged compared to the initial systems in Fig. 4-15, while the elements whose glass has been substituted show much smaller autofluorescence contribution.

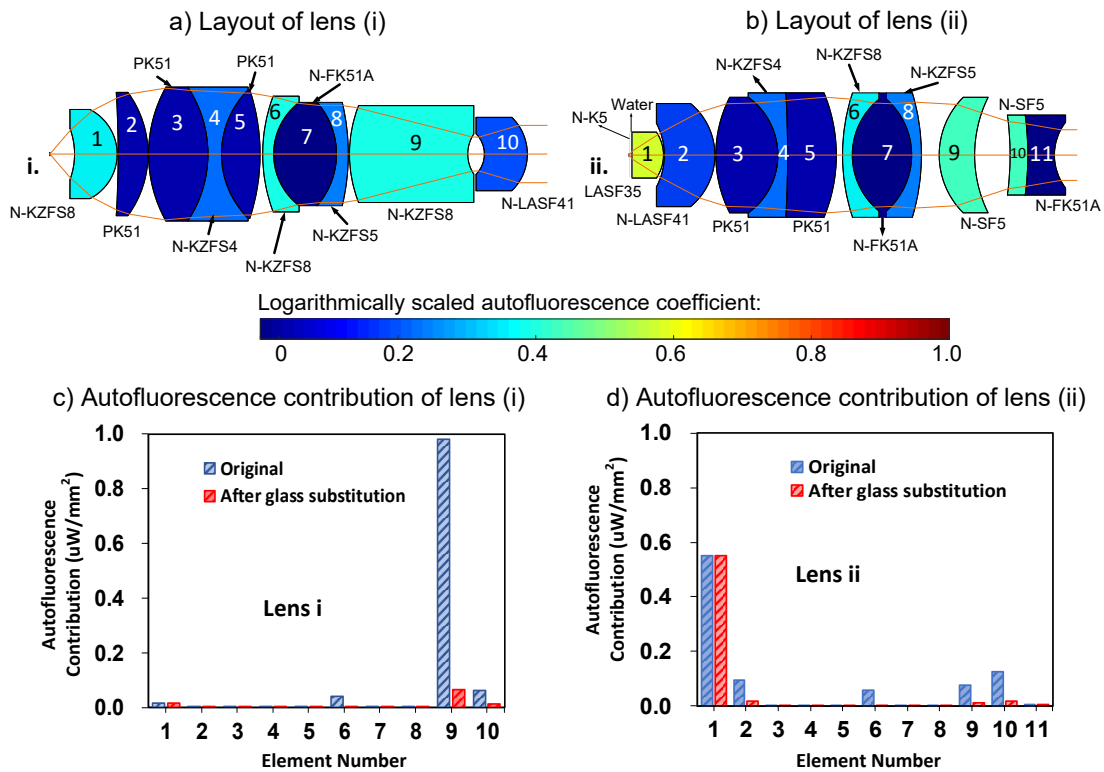


Fig. 4-21. Layouts and autofluorescence contribution of lenses (i) and (ii) after glass substitution. The light source and normalization are the same as described in Fig. 4-15.

The overall autofluorescence intensity of lens (i) is decreased by 90% while that of lens (ii) is decreased by 32%. From Fig. 4-21(d) we can see that the autofluorescence contribution from element 1 of lens (ii) remain unchanged because no substitution glass can be found for the glass LASF35, and the lack of suitable substitution for this large refractive index glass is the major limitation for further improvement of lens (ii).

Based on the above discussion, it is clear that glass substitution is capable of effectively suppressing the autofluorescence contribution with only minor effort, but on the other hand it is limited for critical lens groups, in which special glass types are used.

4.2.4.2 Lens splitting

In order to overcome the limitations of glass substitution, modification of lens structure is necessary to further reduce the autofluorescence intensity, especially for the critical lens groups. In Secs. 4.2.3.1 and 4.2.3.2, we have found that the autofluorescence contribution from different parts of critical lens groups strongly varies, and most of the autofluorescence is contributed by a small part of the element near the intermediate image (see Fig. 4-17 and Fig. 4-19). This means that if we only substitute the glass in this

small part of the critical element, it is possible to effectively reduce the autofluorescence intensity without sacrificing the system performance.

Based on this idea, we split the first element of lens (ii) and element 9 of lens (i) into cemented doublets and substitute the glass of the smaller part with N-FK51A, which is a crown glass with extremely low autofluorescence coefficient. As shown in Fig. 4-22 and Fig. 4-23, after the lens splitting, the radii of curvatures of the cementing surfaces are further adjusted to restore the aberration correction functionalities of the splitted lens groups.

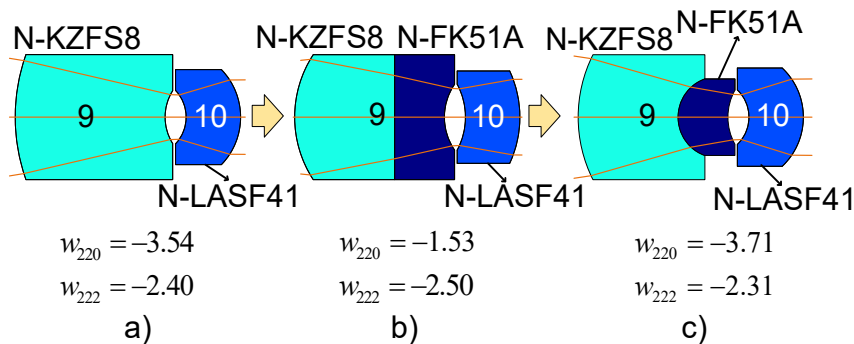


Fig. 4-22. Splitting of the rear group of lens (i). a) shows the original structure of the rear group, b) shows the structure with a planar cementing surface and c) shows the structure with a curved cementing surface. w_{220} and w_{222} are the Seidel coefficients for Petzval curvature and astigmatism in waves, here it represents the overall Petzval curvature and astigmatism contribution of the rear group. From the values of the Seidel coefficients, it is obvious that the aberration contributions from the initial (a) and final (c) systems are very similar, while the Petzval curvature contribution of (b) greatly differs from that of the initial system.

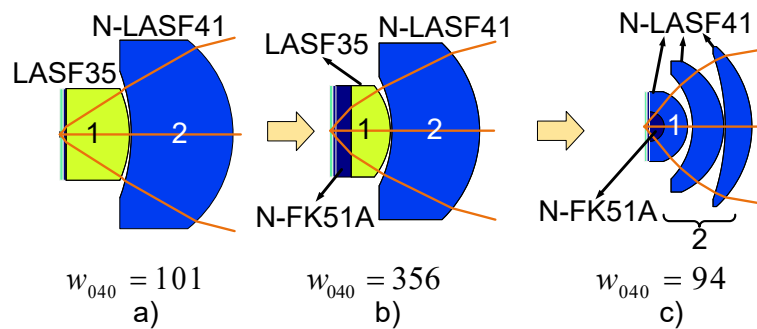


Fig. 4-23. Splitting of the front group of lens (ii), the two layers in the front of element 1 are the cover glass and immersion medium. a) shows the original structure of the front group, b) shows the structure with a planar cementing surface in element 1 and thick meniscus lens in element 2, c) shows the structure with a curved cementing surface in element 1 and thin shell lenses in element 2. w_{040} is the Seidel coefficient for spherical aberration in waves, here it represents the overall spherical aberration contribution of the front group.

Fig. 4-24 shows the layouts of the lenses (i) and (ii) after glass substitution and lens splitting as well as their autofluorescence contributions. According to Fig. 4-24(c,d), after glass substitution and lens splitting, the overall autofluorescence intensity is reduced by 95% for lens (i), and 80% for lens (ii).

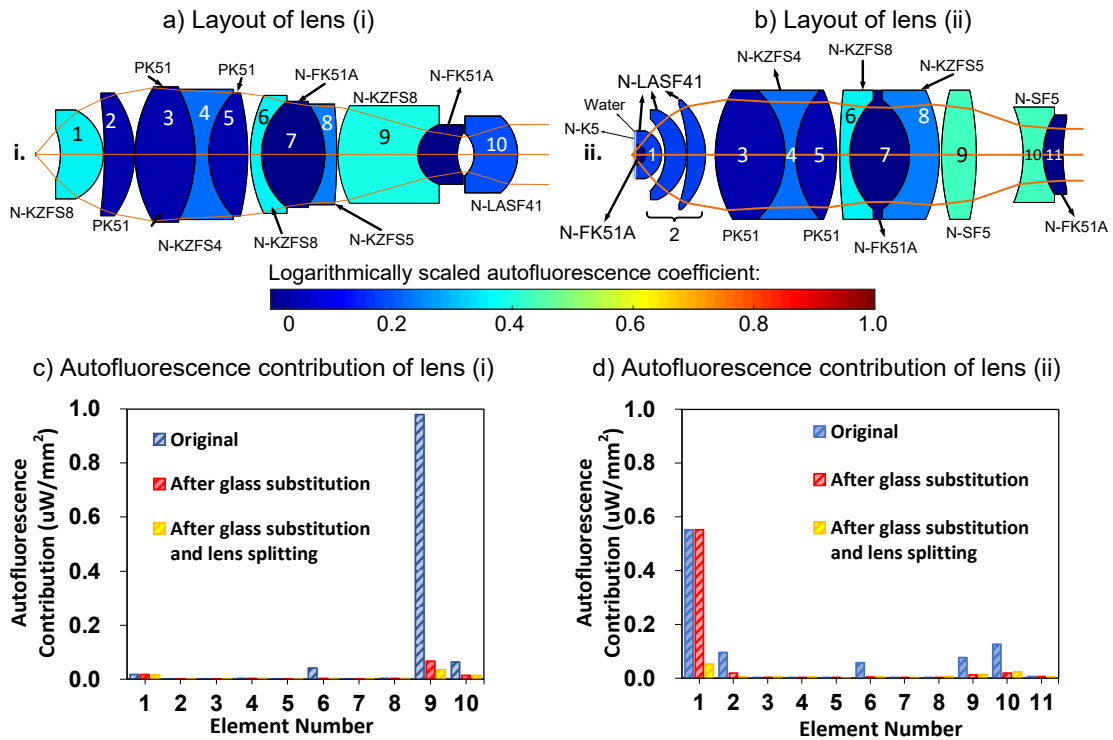


Fig. 4-24. Layouts and autofluorescence contribution of lenses (i) and (ii) after glass substitution and lens splitting. The light source and normalization are the same as described in Fig. 4-15.

With the autofluorescence intensity of lens (i) and (ii) significantly suppressed by the abovementioned methods, the autofluorescence contribution from the cover glass becomes dominant again. For lens (ii), the autofluorescence contribution from the low-fluorescent N-FK51A cover glass is one time larger than that from the lens itself. Because the glass N-FK51A is already one of the most low-fluorescent glass, further reduction of the autofluorescence contribution from the cover glass is not possible from the material point of view, which means that the only option is to reduce the thickness of the cover glass, or even avoid using cover glass for fluorescence microscopy.

4.2.4.3 Optimization strategies

In the previous section, the methods to reduce the autofluorescence intensity of existing lens designs have been discussed. These methods modify the critical lens elements and reduces the autofluorescence intensity by structural and material change. However, considering that the dependence of autofluorescence intensity on the lens structure has been revealed by the quasi-analytical model, and the efficient phase-space-based code enables real-time calculation of autofluorescence contribution, it is more efficient to integrate the autofluorescence calculation starting from the early phase of microscope lens design. In order to do so, judicious optimization strategies are necessary to find the solution which is optimal for both the aberration correction and autofluorescence reduction. Here we discuss the possible approaches to integrate autofluorescence calculation in the optimization of microscope lenses.

Considering that the autofluorescence coefficients of glass are deterministic for the autofluorescence contribution from an element, we can restrict the glass map to the low-fluorescent ‘N-type’ glass from the start of the lens design. However, the drawback of this approach is that the ‘N-type’ glass is not always sufficient to reduce the autofluorescence contribution from critical lens groups, as discussed in Sec. 4.2.4.1. Furthermore, the restricted glass map eliminates the possibility to use highly fluorescent glass for aberration correction in the middle group, which has low geometrical sensitivity to autofluorescence but large contribution to aberration correction. Therefore, different restriction criteria should be applied to different lens groups according to their geometrical sensitivity. However, considering that the geometrical sensitivity of a lens group is strongly dependent on its structure, the determination of such restriction criteria is a non-trivial task.

Compared to the restriction of the glass map, a smarter approach is to integrate autofluorescence in the merit function by adding penalty terms according to the autofluorescence calculation. These penalty terms should reflect the real autofluorescence contribution of each lens element, and larger autofluorescence contribution leads to larger merit function values. In order to accelerate the optimization algorithm, the quasi-analytical model according to Eq. (4.2) can be used for the fast estimation of autofluorescence contribution from the lens elements. The weighting of such penalty terms should be tuned carefully to minimize the autofluorescence intensity without large sacrifice of the aberration correction. After a local minimum has been found, glass substitution can be performed by Hammer optimization to search for better glass combinations.

5 Calculation of Surface Scattering

In sections 3.3 and 3.6, we have discussed the geometrical and physical implementation of the phase space model in the simulation of surface scattering. In this chapter, we demonstrate the applications of the phase space methods in the surface scattering simulation of real optical systems and discuss the applicability and limitations of the method based on the simulation results. For the geometrical implementation of the phase space method, the intermediate image method is first applied to calculate the point source transmittance (PST) of a circularly symmetric Ritchey-Chrétien telescope, and subsequently the reverse raytracing and the hybrid method are applied for the straylight analysis of a Kirkpatrick-Baez telescope with extreme asymmetry.

For the physical implementation of the phase space method based on the Wigner function, the method is applied to calculate the PSFs of a retrofocus lens and an EUV Schwarzschild objective, in which the capabilities of the phase space method to model multiple scattering and diffraction effect are demonstrated.

5.1 Geometrical phase space-based approach

In this section, we present two examples in which the phase space method is applied to calculate the straylight from surface scattering. In addition, the accuracy of the phase space method is validated against the results of the brute force methods, while the pros and cons of the different implementations of the phase space model are discussed in a critical review.

5.1.1 Ritchey–Chrétien telescope

In this example we present the surface scattering simulation for a rotationally symmetric Ritchey–Chrétien telescope (RCT) composed of two hyperbolic mirrors and a field lens, as shown in Fig. 5-1.

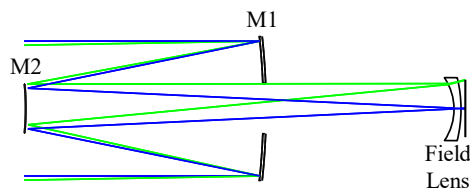


Fig. 5-1. Layout of the Ritchey–Chrétien telescope

Considering the circular symmetry of the RCT, we apply the intermediate image method as shown in Fig. 3-10, so that the impact of all primary aberrations on the detector intermediate image can be represented by the Seidel sums. For subareas on M2, the dominant aberrations in the detector intermediate image are distortion and field curvature

induced by the field lens, while for subareas on M1 the additional coma contributed by M2 also plays an important role.

As the first step of the simulation, we divide M1 and M2 into 289 subareas on an isoenergetic grid with 17 rings, as shown in Fig. 5-2, from which we can see that the irradiance is almost uniformly distributed on the two mirrors except for the central obscuration.

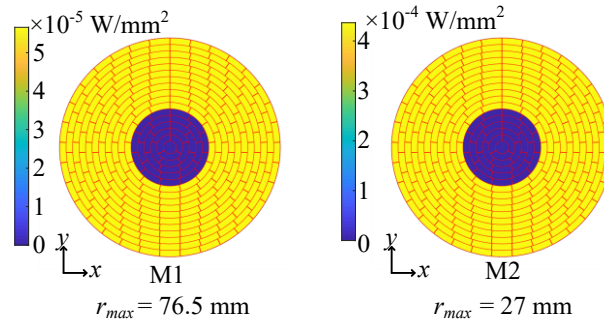


Fig. 5-2. Irradiance distribution on M1 and M2 for an incident beam of 1W. The red mesh indicates the isoenergetic grid on which M1 and M2 are discretized.

The next step of the simulation is to trace rays from the image plane to the target surfaces to determine the intermediate image of the detector seen by the subareas on M1 and M2. In this process, a paraxial marginal ray is traced to determine the location of the intermediate image, after which the paraxial image heights of the pixel centers can be obtained by tracing paraxial chief rays from every pixel center to the intermediate image plane or more conveniently by scaling the detector grid with the magnification.

After obtaining the paraxial intermediate detector image, the shifts of the paraxial pixel points are calculated as described in Sec. 3.3.2.1. The relative positions between the surface subareas and detector pixels determine the field and pupil coordinates, which in turn define the transverse aberrations, from which the pixel shifts are calculated. Additionally, the actual areas of pixels in the intermediate image are calculated based on the distorted detector grid.

The final step of the simulation is to calculate the coupling coefficients between the surface subareas and pixels by Eq. (3.11). In order to do so, the scattering angles, the BRDF values, and solid acceptance angles need to be calculated for every subarea-pixel combination in the corresponding local coordinate systems. The calculation seems to be rather complicated because there are many subarea-pixel combinations, but the algorithm can be vectorized to accelerate the calculations. As an example, Fig. 5-3 shows the coupling coefficients of two tangential subareas on M1 and M2 with the pixels in the corresponding detector intermediate image, where the coupling coefficients indicate the fraction of the incident flux that is scattered to the pixels. From Fig. 5-3 we can see that the peak of the coupling coefficient coincides with the specular reflection, and we can

observe the distorted edges of the intermediate detector image due to aberrations of the field lens and M2.

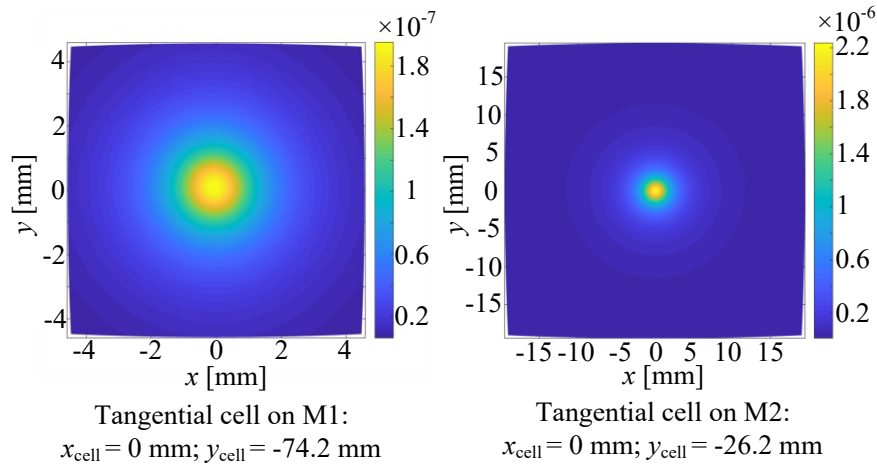


Fig. 5-3. Coupling coefficients between two tangential subareas on M1 and M2 with the intermediate images of the pixels. The x and y coordinates correspond to the positions of the pixel centers in the intermediate images.

After the coupling coefficients for all subareas and pixels are calculated, the PST is obtained by summing up the contributions from all subareas. Here we consider a wavelength of 632 nm and assume M1 and M2 to be smooth surfaces ($\sigma_s = 0.0015 \mu\text{m}$), which results in a small total integrated scattering (TIS) of 0.09 %. Additionally, for the K-correlation model, the slope factor of the PSD is set as $s = 2$, and the correlation length of the surface error is $30 \mu\text{m}$, meaning that small-angle scattering is dominant. Based on these surface parameters, the PSTs of M1 and M2 are calculated and shown in Fig. 5-4, from which we can see that the PST of M2 has a sharper peak than that of M1. This is because the detector intermediate image seen by M2 is nearly four times as large as that seen by M1 and thus the acceptance angles of the pixels are much larger for M2. For the same reason, the total scattered flux from M2 to the detector is twice as large as that from M1, indicating that the M2 is more critical for stray light. This corresponds to the conclusion that elements with smaller marginal ray heights are more critical for stray light.

To validate the simulation results, we have also calculated the PSTs with the Monte Carlo method modelled by non-sequential raytracing in OpticStudio. In order to implement the BSDF modelled by the K-correlation model and the GHS, an external DLL file has been written, which is called by OpticStudio to generate scattered rays.

Fig. 5-5 shows the 1D cross-sections of the PSTs of M1 and M2 calculated by the phase space model and Monte Carlo raytracing. Here we can observe that the signal to noise ratio (SNR) of the result calculated by the Monte Carlo method is much lower than that of the phase space model due to inadequate number of rays that have been traced. Apart from that, a good agreement between the results of the two methods can be observed. On the other hand, the greatest difference lies in the runtime. For the results shown in Fig. 5-5, the Monte Carlo method traces 2.8×10^{10} rays and takes 3.4 hours, while the phase space

model only traces 1×10^5 rays and takes 5.5s to calculate the PSTs of both mirrors with an even higher SNR. Therefore, the phase space model achieved an acceleration ratio of more than 2000 compared to the Monte Carlo method. Furthermore, if the optical system contains more surfaces, or if finer discretization of the detector is required, the advantage of the phase space model in speed and accuracy is more pronounced. The computer platform used for the simulation has two Intel Xeon E5-2690 v4 CPUs (16 cores @ 2.60 GHz).

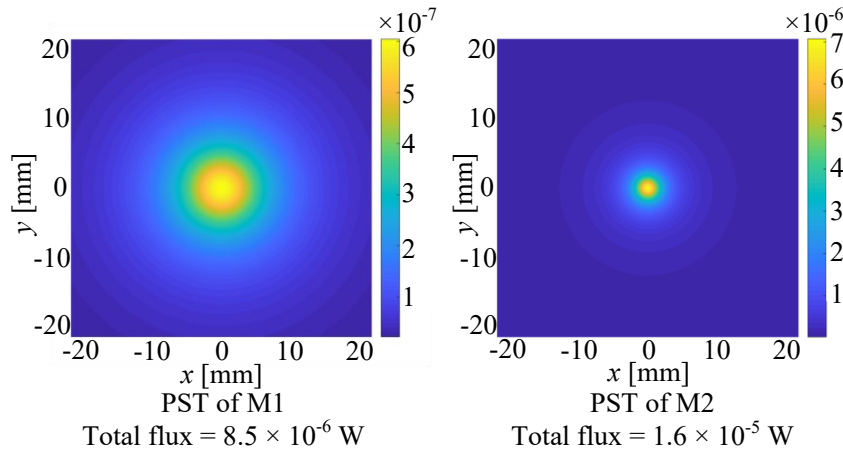


Fig. 5-4. Point source transmittance (PST) function of M1 and M2 for an incident beam of 1 W. The x and y coordinates correspond to the pixel centers on the 100×100 detector. The color code indicates the scattered irradiance on the detector, the units are W/mm^2 .

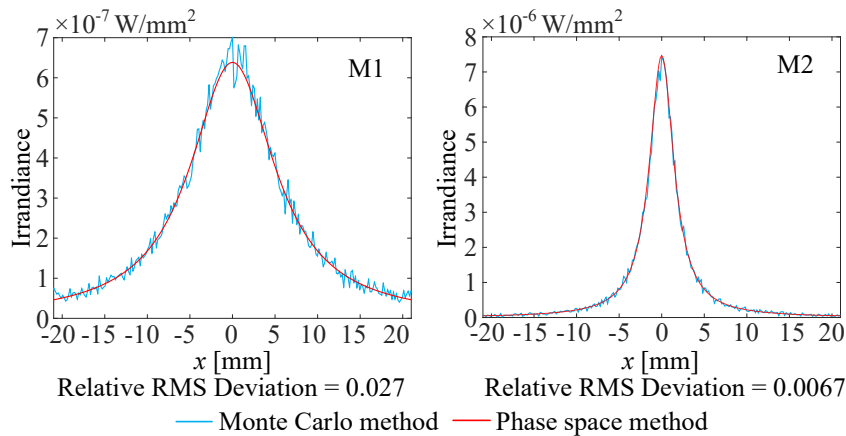


Fig. 5-5. x cross-sections of the PSTs calculated by the phase space model and the Monte Carlo method. The relative RMS deviation is normalized to the peak values in each figure.

5.1.2 Kirkpatrick-Baez telescope

The Kirkpatrick-Baez telescope (KBT) design is widely used in X-Ray optics which requires grazing incidence to obtain high reflectivity. As illustrated in Fig. 5-6, a KBT is composed of two mirrors placed perpendicular with each other. As it can be seen in Table 5-1, the two mirrors are parts of two biconic surfaces, which have focal power only in the y and x directions respectively. Furthermore, the incident angles of the rays on the two mirrors are extremely large. Therefore, circular symmetry is completely broken and

both biconic surfaces introduce significant amounts of astigmatism, making the intermediate image method difficult to apply. Consequently, here we use the inverse raytracing method and the hybrid method to calculate the total scattered flux and the PSTs of M1 and M2.

The surface scattering model used in this example is the same as that used for the RCT. Again, we apply the GHS theory to predict the BRDF due to its accuracy for large incident and scattering angles. Here an incident beam with a short wavelength of 10 nm is considered, and the correlation length of surface error is 100 nm. Therefore, wide-angle scattering from high spatial frequency surface roughness is dominant.

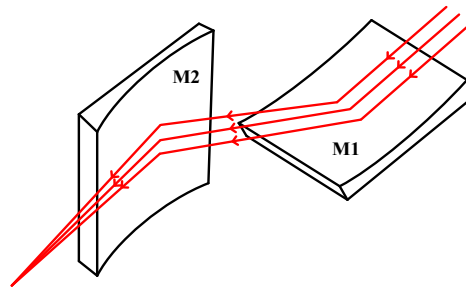


Fig. 5-6. Layout of the Kirkpatrick-Baez telescope

Table 5-1. Surface data of the Kirkpatrick-Baez telescope

Surface	x -radius	x -conic	y -radius	y -conic	Incident angle of axis ray
M1	infinity	0	-1 mm	-1	82.5°
M2	0.209 mm	-1	infinity	0	77.6°

Due to the loss of circular symmetry, a Cartesian grid with 80×80 sampling points is used for surface discretization. The entrance pupil is sampled by a Fibonacci grid with 10^5 points, from which the incident rays are traced to the two mirrors. As shown in Fig. 5-7, the forward raytracing results give the irradiance distribution and the incident angles in each surface subarea.

Subsequently, four sets of rays are traced from the corners of the detector back into the optical system, and the solid angle between the rays from the four vertices gives the solid acceptance angle $d\Omega$ for each subarea of the two mirrors, as shown in Fig. 5-7(c, d). From these two figures we can see that the value of $d\Omega$ is not uniformly distributed on each mirror, which mainly results from the large astigmatism introduced by the cylindrical surfaces and the projection effect due to grazing incidence. It is also evident that for both mirrors $d\Omega$ is rather small due to the small FOV, but here the values of BRDF cannot be assumed to be constant within $d\Omega$ because the BRDF is very sensitive to the scattering angle under grazing incidence. Therefore, the exact integration according to Eq. (3.9) must be calculated by numerical methods.

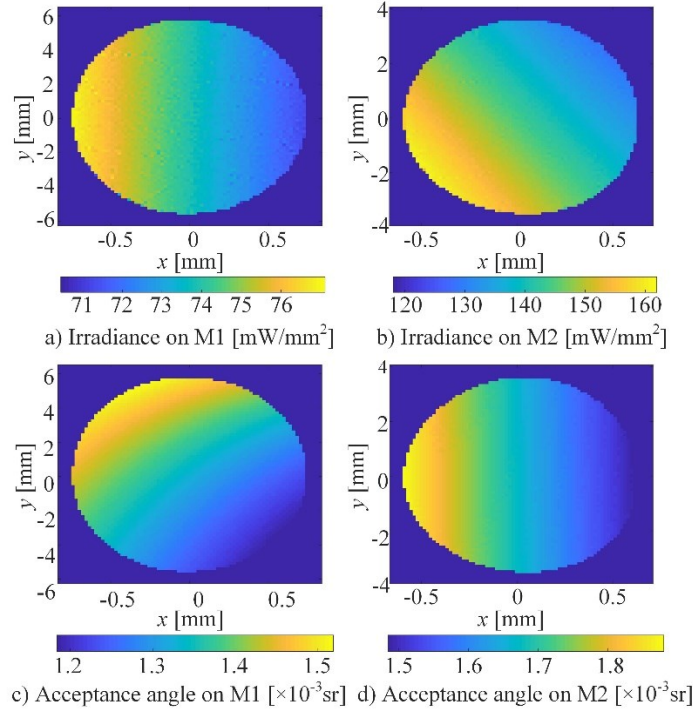


Fig. 5-7. Irradiance and solid acceptance angle distribution on M1 and M2 in local coordinate systems.

Finally, by summing up the scattered flux from all the subareas of each mirror, we obtain the contribution of each mirror to the straylight on the detector, as shown in Fig. 5-8. The large contribution from M2 mainly results from its larger solid acceptance angle. To validate these results, the Monte Carlo method is again used for comparison, as indicated by the orange bars in Fig. 5-8, from which we observe a good agreement between the two methods. The overall calculation time of the phase space method is 48 s which includes the tracing of 5×10^5 rays and data processing, while the Monte Carlo method traces 8×10^8 rays and takes 720 s. Here the acceleration ratio is not as large as that in Sec. 3.1 due to two reasons, the first being that processing the raytracing data is time-consuming when fine surface sampling is used, and the second being that only a small number of rays need to be traced in the Monte Carlo method since only the total scattered flux is calculated. However, here we have used a rather fine sampling grid to generate the figures with higher resolution for clearer demonstration, while in practice a coarser sampling of 10×10 is enough to generate similar results, as shown by the green bars in Fig. 5-8. By comparing the green and blue bars in Fig. 5-8, we find that the coarse sampling grid yields virtually identical results to the dense sampling grid, while shortening the runtime to 3 s.

Additionally, as shown in Fig. 5-9, the PSTs of the two mirrors are calculated by the hybrid method. As a validation of the hybrid method, Fig. 5-10 shows the PSTs calculated by the Monte Carlo method.

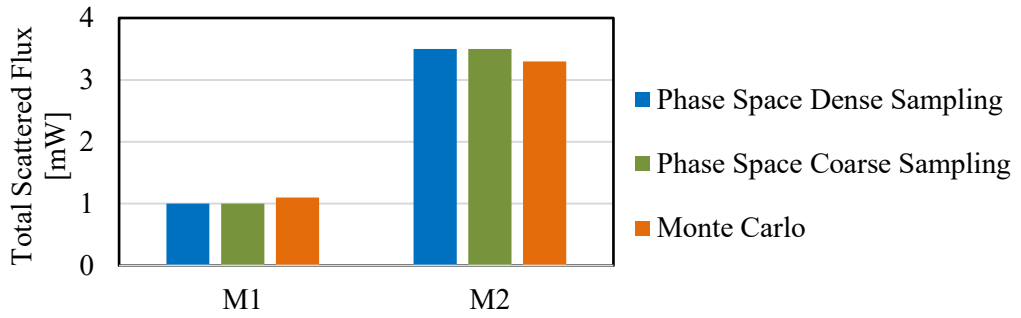


Fig. 5-8. Total scattered flux on the detector contributed by M1 and M2. The results are calculated by the phase space model and Monte Carlo method.

Comparing Fig. 5-9 and Fig. 5-10, we find that the hybrid method yields similar results to the Monte Carlo method, which can also be seen from the cross-sections shown in Fig. 5-10(c, d). The two dark holes in Fig. 5-10(a, b) result from the element used to block the specular rays in the non-sequential raytracing. They are not seen in Fig. 5-9 because the scattered and specular rays can be directly decoupled in the phase space method. Additionally, from the relative RMS deviation of the results shown in Fig. 5-10(c,d), we observe a larger deviation for M1, for which there are mainly two reasons. The first is that the detector acceptance angle on M1 is very small, resulting in low SNR of the result calculated by the Monte Carlo method, and the second reason is the large astigmatism contribution from M2. Since astigmatism is not linearly dependent on the field coordinate, the distortion of the detector grid due to astigmatism cannot be accurately modelled by the linear transformation matrices used in the hybrid method. Therefore, if large aberrations which are not linearly dependent on the field coordinate are present in the detector intermediate image, modelling the detector grid distortion by linear transformations induces errors, which decreases the robustness of the hybrid method.

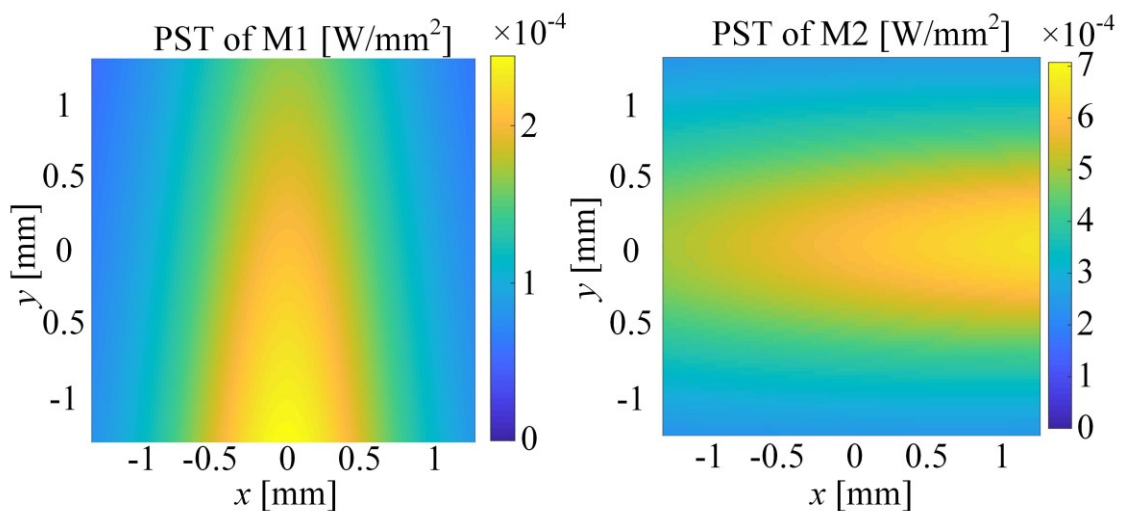


Fig. 5-9. PSTs of the KBT for an incident beam from the on-axis field with a power of 1 W. The detector is a 100×100 square pixel array with a width of 2.6 mm.

Furthermore, from the PSTs shown in Fig. 5-9, we observe that the axes of symmetry of the two PSTs, which correspond to the planes of incidence, are perpendicular to each

other as a consequence of the perpendicular orientations of the two mirrors. Furthermore, we observe that the peaks of the PSTs do not coincide with the specular direction, which results from the fact that the peak of the BRDF does not coincide with the specular reflection for large incident angles as described by the GHS theory [11].

As expected, the hybrid method shows great advantage in terms of runtime. The calculation of the PSTs shown in Fig. 5-9 takes only 7 seconds for both mirrors when using a coarse 5×5 Cartesian grid for surface discretization. For this coarse surface discretization, 50000 rays have been traced to calculate the irradiance and acceptance angle distribution on the optical surfaces. Similar to the inverse ray tracing method shown in Fig. 5-8, increasing the discretization density of the optical surfaces only yield minor increase of the accuracy, while the number of traced rays is linearly dependent on the number of subareas and so is the calculation time.

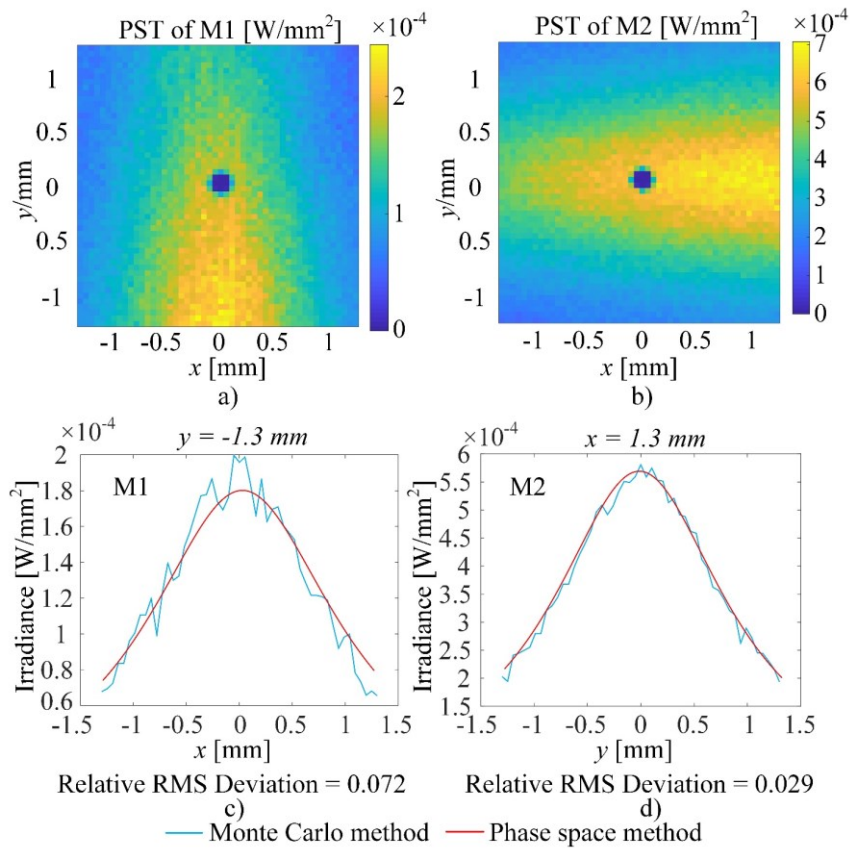


Fig. 5-10. PSTs of the KBT calculated by the Monte Carlo method. The dark holes in the centers of the first two figures result from the element that blocks the specular rays. The bottom figures show the comparison of the x and y cross-sections of the PSTs calculated by the Monte Carlo method and the phase space model. The Relative RMS deviation is normalized to the peak values in each figure.

In contrary, the Monte Carlo method traces 4×10^{10} rays, taking 11 hours to produce a result with a relatively poor SNR (see Fig. 5-10). Therefore, an acceleration factor of 5600 is achieved for this case. Additionally, if the SNR of the PST of M1 is increased to be the same as that of M2, the total calculation time is increased to 22 hours. The speed advantage of the hybrid method is particularly pronounced for this case because the

BRDF is concentrated in a small angular range under grazing incidence, making the random sampling of the BRDF in the Monte Carlo method extremely slow.

5.1.3 Critical review

In the last two sections we have used examples to discuss the principles of the three methods of calculating the straylight level on the image plane based on the phase space model. In this section, we present a critical review and detailed comparison of the three methods. Based on the critical review, we discuss the applicability and limitations of each method as well as the selection criteria for different simulation scenarios.

Table 5-2 shows a detailed comparison of the applicability and limitations of the three methods, from which we can see that the intermediate image method should be used as long as the system is circularly symmetric, while for systems without circular symmetry, the inverse raytracing method should be used if no detector discretization is required. The hybrid method is the only option if the scattered irradiance distribution on the detector is to be calculated for systems without circular symmetry. Therefore, the hybrid method is currently the most general method that is suitable for nearly all optical systems, while its major weakness being the lack of robustness due to the neglect of the non-linear field-dependent aberrations.

Table 5-2. Comparison between the three methods of the phase space model

	Intermediate image method	Inverse real ray-tracing method	Hybrid method
Restriction for circular symmetry	Yes	No	No
Robustness	High	High	Low in the presence of aberrations which are not linearly dependent on the field coordinate
PST calculation	Yes	No	Yes
Support for freeform surfaces	No	Yes	Yes
Efficiency for dense surface sampling	High	Low	Low

In general, the phase space model is completely geometric, and diffraction effects are not considered. The scattered light distribution on the detector is obtained by incoherently superposing the scattered light from different surface subareas. Therefore, the phase space model is expected to yield accurate results as long as the correlation area of the residual surface error is smaller than the area of the surface subareas, so that the scattered light from different subareas can be considered to be incoherent. On the other hand, the Monte Carlo method assumes an infinitely small correlation area of the residual surface error, which means that it can only model scattering due to high frequency surface roughness, while the phase space model is able to model scattering from mid-spatial-frequency (MSF) errors by selecting a proper size for the surface subareas.

As discussed in the last two sections, the phase space model shows great advantage over the Monte Carlo method in terms of accuracy and runtime. Due to the parallelization of the subarea-pixel coupling algorithm, the intermediate image method is the most efficient, while the inverse raytracing and hybrid methods are less efficient because processing the reverse raytracing data is time-consuming. In order to evaluate the relationship between the accuracy of the phase space model with the sampling density and runtime, we use an example of a single lens imaging system with a large FOV shown in Fig. 5-11(a) and calculate the PST of the front surface by the intermediate image method, the hybrid method, and the Monte Carlo method.

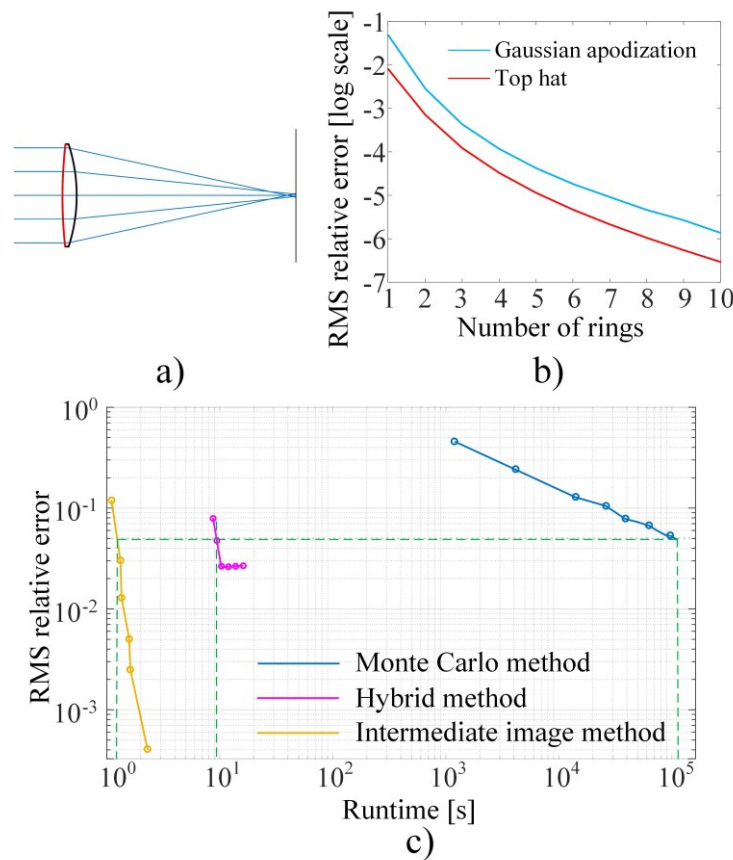


Fig. 5-11. a) Layout of a single lens imaging system. b) RMS relative error of the simulation results vs. the number of rings in the isoenergetic polar sampling grid. c) Log-log plot of the RMS relative error of the simulation results vs. runtime. The errors are determined relative to the results calculated by the intermediate image method with 20 rings.

The logarithmically scaled RMS relative error of the result calculated by the intermediate image method with different numbers of rings used for surface discretization is depicted in Fig. 5-11(b). From this figure we can see that the relative error rapidly converges to very small values as the number of rings increases, and the relative error is larger in the presence of apodizations, meaning that more rings are needed if the irradiance distribution on the optical surface is not uniform. Fig. 5-11(c) shows the dependence of the RMS relative error on the runtime. Large runtime differences between the methods leads to significant separation of their curves, and here the phase space model holds the clear

advantage. In Fig. 5-11(c), from left to right, the runtime is increased due to the increase of the sampling density of the optical surface or the increase of the number of traced rays. From Fig. 5-11(c), we see that the intermediate image method is the most efficient and its error converges rapidly to zero, while the hybrid method is slower than the intermediate image method but still much faster than the Monte Carlo method. It should also be noted that the error of the hybrid method converges to a finite value instead of zero, which results from the fact that the large distortion, astigmatism, and Petzval curvature introduced by the rear surface of the lens due to the large FOV cannot be accurately modelled by the linear transformations used in the hybrid method. As indicated by the green dashed line in Fig. 5-11(c), for an RMS relative error of 0.05, the intermediate image method achieves an acceleration ratio of 10^5 compared to the Monte Carlo method. Here the acceleration factor is strongly dependent on the required RMS relative error and is expected to be even higher if smaller RMS relative error is required.

5.2 Wigner function-based simulation of surface scattering

The Wigner function-based propagation of partially coherent light in optical systems with scattering surfaces has been introduced in Sec. 3.6, in which the Wigner function method is used to model the scattering of partially coherent light by microroughness and MSF errors based on statistical and deterministic surface models.

In real optical systems, all frequency components of surface errors coexist, and multiple scattering from these surface errors must be taken into account if the optical system is composed of more than one surface. Furthermore, although the field scattered by surface roughness is incoherent, it becomes partially coherent after it has propagated any distance from the scattering surface according to the Van Cittert-Zernike theorem [18]. Therefore, the diffraction of the scattered field from the subsequent edges and MSF structures need to be considered. Since the Wigner function is able to model partially coherent light, the edge diffraction of the scattered and specular light can be simultaneously modelled with surface scattering. In this section, we demonstrate the Wigner function-based simulation of multiple scattering from MSF surface errors and surface roughness in a retrofocus lens and a Schwarzschild EUV objective.

5.2.1 Retrofocus lens

We consider a simplified two-element retrofocus lens composed of a negative lens followed by a positive lens, as shown in Fig. 5-12. The incident Gaussian-Schell beam is truncated by the front element and the stop. The distances between the stop and the two lenses are 15mm. Both lens surfaces possess the same MSF structure with a period of $T = 120 \mu\text{m}$ and a PV value of $h_{pv} = 0.5 \mu\text{m}$, and the standard deviations of the tool tip displacement in the thrust and feed directions are $\sigma_t = 0.05 \mu\text{m}$ and $\sigma_f = 4 \mu\text{m}$. Additionally,

the surface roughness of the two lens surfaces is also the same, with a correlation length of $l_c = 20 \mu\text{m}$ and a RMS surface height of $\sigma_s = 0.1 \mu\text{m}$. The incident beam is a top-hat Gaussian Schell beam with a beam width of 7 mm and a wavelength of 1 μm .

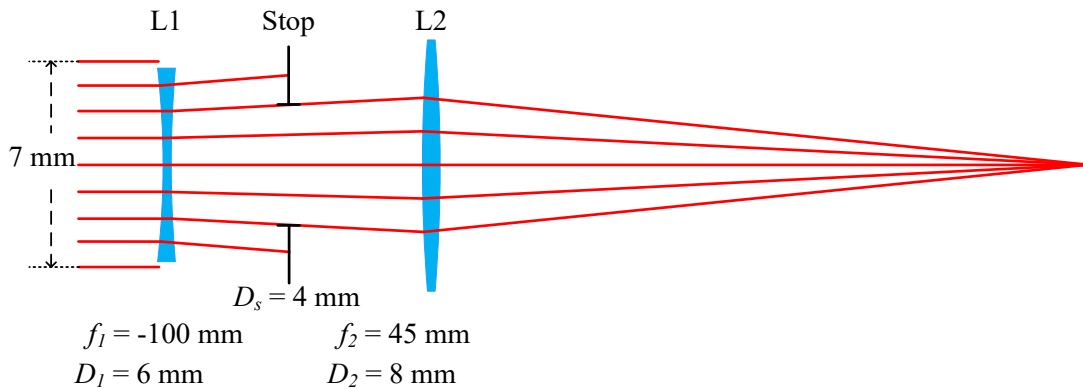


Fig. 5-12. A simplified two-element retrofocus lens. The incident Gaussian-Schell beam is truncated by the front element and the stop. The distances between the stop and the two lenses are 15mm.

The simulation results are shown in Fig. 5-13 and Fig. 5-14. In Fig. 5-14(a) we can observe the beamlets formed by the MSF structures, as well as a bright background of the beam after L2, which results from the wide-angle scattering induced by the surface roughness. From the magnified figure in Fig. 5-14, we can observe that the scattered and diffracted field from L1 and the stop, which is the part of the beam outside the specular region, is scattered again by L2, forming the beamlet structure. The PSF of the system are depicted in Fig. 5-14(b), from which we observe the sidelobes due to the MSF structures and a non-zero offset resulting from the surface roughness. Furthermore, the cascaded diffraction from the edges of L1 and the stop is also considered in the simulation. Fig. 5-13(c) shows the Wigner distribution of the beam after the stop, from the enlarged figure of the Wigner distribution near the edge of the stop we can observe the ripples induced by edge diffraction. Therefore, the Wigner function-based simulation is able to simultaneously model multiple scattering from MSF errors and surface roughness, as well as diffraction from lens and stop edges.

The Wigner function approach is much more efficient than the classical straylight analysis methods due to the absence of stochastic raytracing. For the above-mentioned simulation of the retrofocus lens, we have applied Radon transform instead of shearing to model the free space propagation after L2, which further accelerate the calculation and eliminates the aliasing effect due to the spreading of the scattered field outside of the x -limit [33]. For this simulation, the runtime of the PSF calculation without the intermediate z -planes is 60s for a Wigner function resolution of 2048×2048 based on a PC with Intel Core i5 CPU (6 Cores @ 3.3GHz).

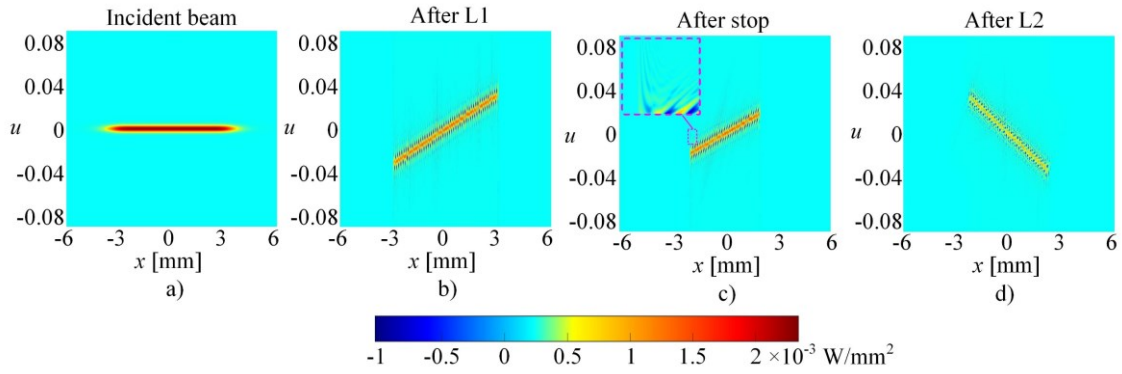


Fig. 5-13. Wigner distributions of the beam in the retrofocus lens.

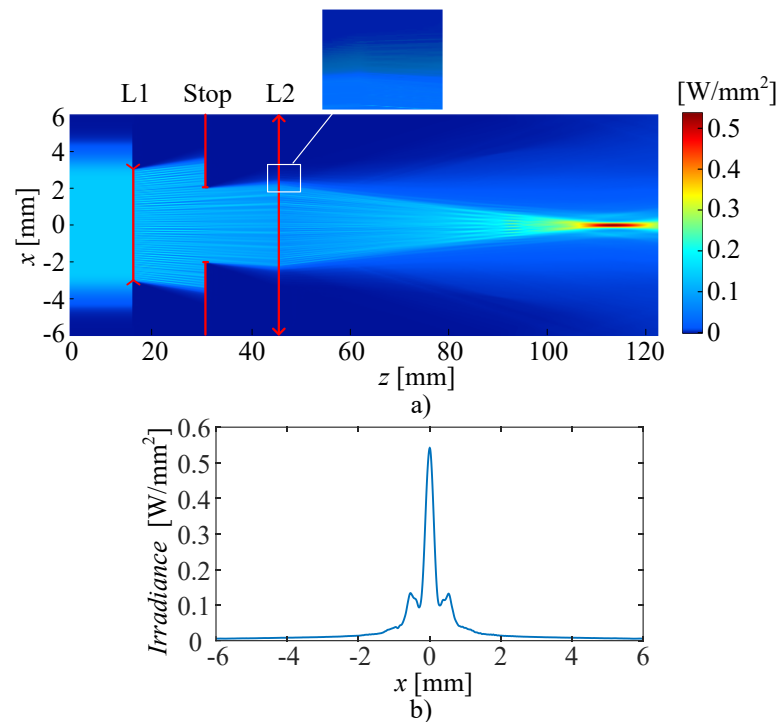


Fig. 5-14. a) Irradiance distribution of the beam in the retrofocus lens. b) the x -cross-section of the irradiance distribution at the beam focus.

5.2.2 EUV Schwarzschild objective

Considering that the TIS is related to the effective surface roughness, which is scaled by the wavelength, surface scattering has particularly large impact on the image quality of mirror systems that work with short wavelengths. To demonstrate the application of the Wigner function method for the straylight analysis of such mirror systems, as shown in Fig. 5-15, we consider an EUV Schwarzschild objective [34] composed of two mirrors, on which MSF errors and micro roughness coexist. For simplicity, we consider the 1-dimensional case.

In this example, we assume that both mirror surfaces of the Schwarzschild objective possess the same MSF structure with a period of $T = 200 \mu\text{m}$ and a PV value of $h_{pv} = 2.5 \text{ nm}$, and the standard deviations of the tool tip displacement in the thrust and feed

directions are $\sigma_t = 0.3$ nm and $\sigma_f = 5$ μm . Additionally, the microroughness properties of the two lens surfaces are identical, with a correlation length of $l_c = 27$ nm and an RMS surface height of $\sigma_s = 0.5$ nm. The partially coherent EUV incident beam is modelled by a top-hat Gaussian Schell beam in addition to a parabolic wavefront.

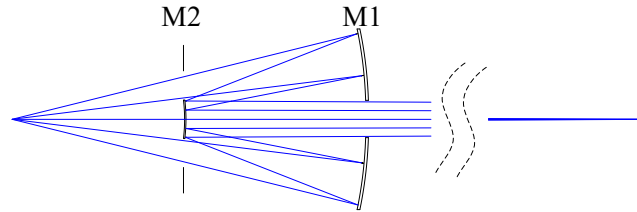


Fig. 5-15. An EUV Schwarzschild objective with an object space numerical aperture (NA) of 0.19 and an image space NA of 0.004. Only part of the ray bundle in the image space are shown due to visualization purposes. The diameters of M1 and M2 are 50 mm and 10.6 mm respectively.

As we can see in Fig. 5-15, the two mirrors are much larger than the optical components discussed in Sec. 3.6.2, and the wavelength of the EUV light is much shorter than the NIR light discussed above. Consequently, the maximum angle that can be described by the Wigner function is strongly constrained by the bandwidth limit of the discrete Fourier transform (DFT). For example, the maximum angle that can be described by a Wigner function with 14000×14000 sampling points is only 0.4 degree for this EUV Schwarzschild objective. The small angular range of the Wigner function is not a problem for the description of surface scattering since the MSF surface errors produce small-angle scattering within the angular range of the Wigner function, and the light scattered by microroughness to large scattering angles does not reach the detector. However, the small angular range of the Wigner function prohibits us from describing high-NA beams in the optical system since the marginal ray angles are much larger than the angular limit of the Wigner function. In order to solve this problem, we apply a special propagation method to remove the parabolic wavefront from the high-NA beams during propagation, which allows us to convert the convergent and divergent beams into a quasi-collimated beam [33, 35]. In this case, the angular range of the Wigner function only needs to cover the angular spreading induced by surface scattering and the partial coherence of the beam. In this simulation, the aberrations of the system are modelled by a phase plate at the exit pupil, which contains the residual wavefront aberration of the system at the exit pupil represented by the polynomials.

The irradiance distribution of the beam during propagation is shown in Fig. 5-16, in which the mirrors have been unfolded to separate the beams for clear demonstration. From Fig. 5-16(a) we observe the beamlets formed by the MSF structures, as well as a bright background after M1 induced by microroughness scattering. The irradiance in the white area of Fig. 5-16(a) has no influence on the scattered field on the image plane and is not calculated in order to reduce memory consumption. The scattered field after M1 is propagated to M2, by which it is reflected and scattered again. The multiply scattered

field is then propagated back and truncated by the central hole of M1. As a final step, the truncated field is propagated to the image plane, as shown in Fig. 5-16(b).

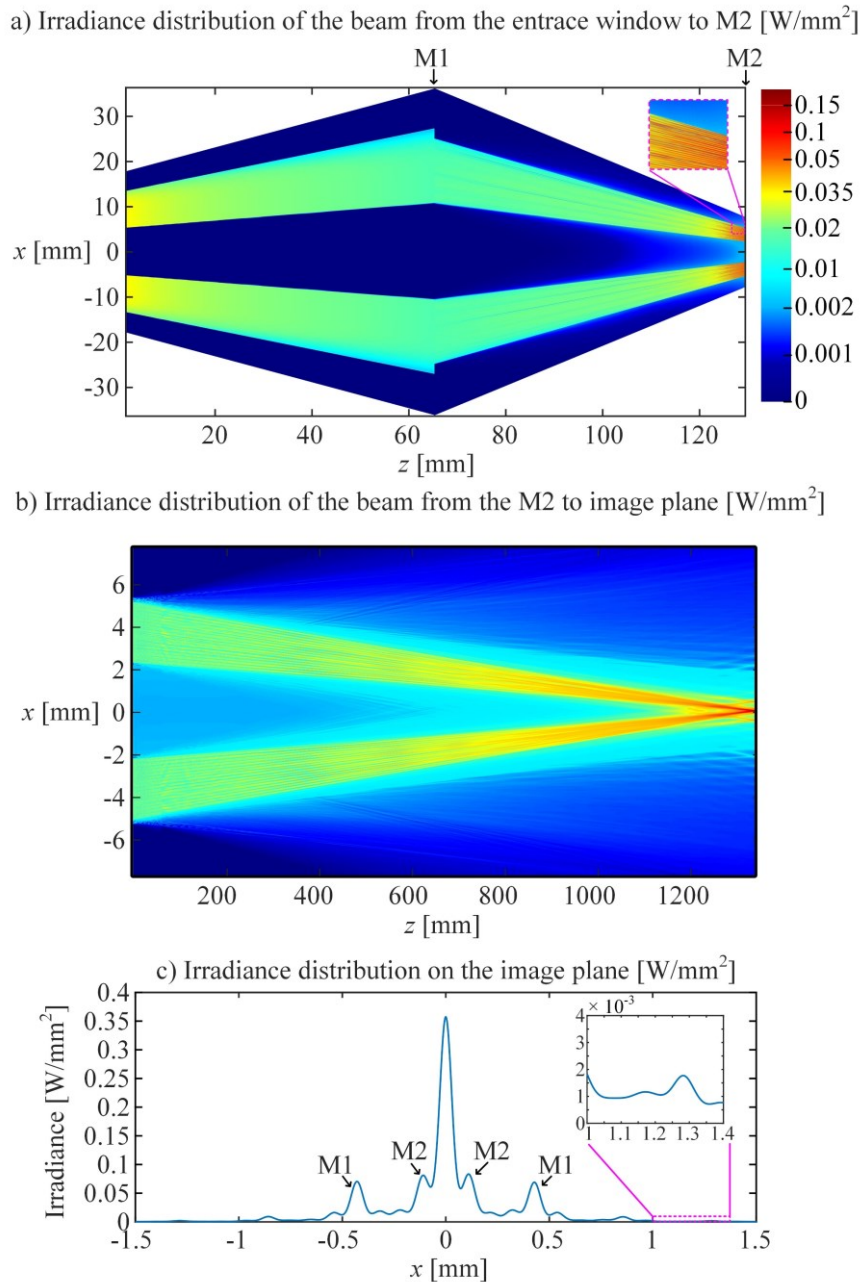


Fig. 5-16. a) 1-dimensional Irradiance distribution of the beam from the entrance window to M1 and M2. b) 1-dimensional Irradiance distribution from M2 to the image plane. c) 1-dimensional irradiance distribution in the image plane.

Fig. 5-16(c) shows the PSF of the system, from which we observe several side lobes around the central peak due to the MSF errors and a small non-zero offset due to the microroughness. By simulating the single scattering from M1 and M2 individually, we can determine that the side lobes near the central peak are generated by the MSF error on M2, while the side lobes further from the central peak originate from the MSF structure of M1. Since the periods of the MSF structures on M1 and M2 are the same, the angular extension of the scattered light from the two mirrors should also be identical. The different

locations of the side lobes originate from the demagnification between M1 and M2, which leads to smaller periods of the phase modulations from M1 at the exit pupil, and thereby larger lobe separation.

The Wigner distributions of the quasi-collimated beam (with parabolic wavefronts removed) after the entrance window, M1, and M2 are shown in Fig. 5-17. From Fig. 5-17(a), we can observe the ripples generated by the edge diffraction at the entrance window, while in Fig. 5-17(b) we observe the complicated structure induced by surface scattering and edge diffraction at M1. Additionally, we observe in Fig. 5-17(c) that the structure of the Wigner function after M2 becomes even more complicated due to multiple scattering. This demonstrates the ability of the Wigner function-based simulation to simultaneously model multiple scattering from MSF errors and surface roughness, as well as diffraction of the specular and scattered light from the edges of the mirrors and the stop.

In order to accelerate the simulation and reduce memory consumption, we have applied various methods including parallelization and parabolic wavefront removal. Additionally, we have applied the Radon transform instead of shearing to model the free space propagation after M2, which further accelerates the calculation and eliminates the aliasing effect due to the spreading of the scattered field outside of the x -limit [33]. For this simulation, the calculation time of the PSF is 170 seconds for a Wigner function resolution of 14336×14336 based on the same computer as in the last section.

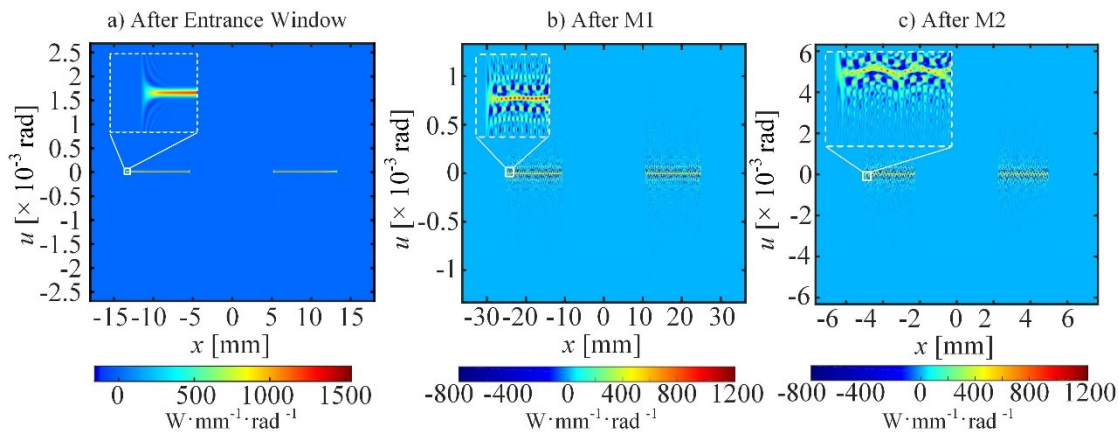


Fig. 5-17. Wigner distributions of the beam in the Schwarzschild objective after the entrance window (a), M1 (b), and M2 (c). The parabolic wavefronts of the beams have been removed.

5.2.3 Limitations of the Wigner function method

Although the Wigner function approach shows great advantage for modelling multiple scattering in optical systems, it also has clear limitations resulting from the paraxial approximation and TEA, which prevent us from simulating high-NA optical systems with the Wigner function.

According to the TEA, Eq. (3.25) is only valid for one incidence angle, and phase modulation of a scattering surface on the incident field is dependent on the incident and scattering angles. Therefore, when the angular spreading of the beam is greatly increased by a scattering surface, applying the TEA to model the MSF errors of subsequent scattering surfaces under the assumption of normal incidence will induce errors. One solution to this problem is to decompose the scattered field into mutually incoherent modes and to propagate each mode through the scattering surface. However, mode expansion greatly increases the computational effort as the number of scattering surface increases. On the other hand, this problem is less critical for aperture-dominant systems in which only the scattered light inside a small angular range around the specular ray can reach the detector, allowing large scattering angles to be ignored.

Additionally, it should be noted that the Wigner function-based beam propagation method is completely paraxial and based on the ABCD matrix, including the parabolic wavefront removal method which we have used to propagate the high-NA beams in Sec. 5.2.2. Therefore, errors are unavoidable during the propagation of high-NA beams. Furthermore, in the abovementioned simulations, we have modelled the aberrations of the optical system by a phase plate at the exit pupil, while the real case, the beams inside the optical system is already aberrated. In the optimal case, the aberration contribution of every optical surface should be included in the Wigner function after the surface, and the aberrated beams should be propagated between the optical surfaces, so that the induced aberrations can also be considered. In principle, modelling the aberrations of optical surface is possible with the Wigner function approach, as long as the surface sags can be neglected, and the aberrations can be represented by ‘aberration sheet’. However, for strongly curved surfaces, neglecting the surface sag leads to errors in the aberration calculation. One possible solution to overcome the limitation of surface sags is to generalize the TEA by the local spherical interface approximation (LSIA) [36].

So far, we have limited our simulations in 2D, while simulation in 3D is necessary for real optical systems. In case of 3D beam propagation, the Wigner function becomes a 4D function and thus greatly increases the memory consumption and computational complexity. Therefore, the resolution of the Wigner function is limited in the 3D simulations, which makes it difficult to model large surfaces with many periods in the MSF structures.

Although the Wigner function approach shows great advantage for modelling multiple scattering in optical systems, it also has clear limitations resulting from the paraxial approximation and TEA, which prevent us from simulating high-NA optical systems with the Wigner function.

According to the TEA, Eq. (3.25) is only valid for one incidence and scattering angle, and phase modulation of a scattering surface on the incident field is dependent on the incident

and scattering angles. Therefore, when the angular spreading of the beam is greatly increased by a scattering surface, applying the TEA to model the MSF errors of subsequent scattering surfaces under the assumption of normal incidence will induce errors. One solution to this problem is to decompose the scattered field into mutually incoherent modes and to propagate each mode through the scattering surface. However, mode expansion greatly increases the computational effort as the number of scattering surface increases. On the other hand, this problem is less critical for aperture dominant systems in which only the scattered light inside a small angular range around the specular ray can reach the detector, allowing large scattering angles to be ignored.

Additionally, it should be noted that the Wigner function-based beam propagation method is completely paraxial and based on the ABCD matrix, including the parabolic wavefront removal method which we have used to propagate the high-NA beams in Sec. 5.2.2. Therefore, errors are unavoidable during the propagation of high-NA beams. However, there also exist non-paraxial formulations of the Wigner function [37], as well as the corresponding propagation methods [38], which could be utilized to overcome the paraxial limitation.

Furthermore, although the aberrations induced by the optical surfaces can be modelled by a series of differential operators acting on the Wigner function or by a phase plate that contains all the wavefront aberrations [39-41], doing so requires the neglect of the surface sag and approximating the optical surfaces as thin sheets, which induces errors for optical surfaces whose surface sag cannot be neglected. One possible solution to overcome the limitation on the surface sag is to generalize the TEA by the local spherical interface approximation (LSIA) [36].

So far, we have limited our simulations in 2D, while simulation in 3D is necessary for optical systems without circular symmetry. In case of 3D beam propagation, the Wigner function becomes a 4D function and thus greatly increases the memory consumption and computational complexity. Therefore, the resolution of the Wigner function is limited in the 3D simulations, which makes it difficult to model large surfaces with many periods in the MSF structures.

6 Conclusions

In this work, we have proposed a phase-space-based method to model light scattering in optical systems and presented the implementations of these method in the geometrical and physical simulation of volume and surface scattering. The geometrical phase space method shows great advantage over the Monte Carlo methods in efficiency and accuracy due to the quasi-analytical phase space coupling of the generation and acceptance of scattered light. While the physical implementation of the phase space method based on the Wigner function is made possible by establishing a combined definition of light coherence, which then provides the possibility to model multiple scattering of partially coherent light by the MSF and HSF surface errors together with diffraction and interference.

As an example, we have first implemented the phase space method to simulate the autofluorescence effect of microscope lenses, in which the autofluorescence generation is modelled as volume scattering and the coupling of the fluorescence light to the detector is analyzed in every z -slice of the elements. With this method, we are able to efficiently calculate the contribution of autofluorescence from every element, which allows us to identify the critical elements for straylight in real time during optical design. Based on the simulation results, we have performed a systematic analysis of the autofluorescence performance of a variety of microscope lenses, and we have shown that the phase space method can be effectively integrated in the optical design process to suppress the straylight from volumes scattering.

Furthermore, we have extended the phase space method to the simulation of surface scattering and investigated the possibilities to generalize the phase space method by removing the critical approximations which restricts its application in certain types of optical systems. In order to eliminate the paraxial approximation and take the aberrations into account, we have utilized the Seidel coefficients to model the impact of aberrations on the straylight coupling. While we have implemented real inverse raytracing to overcome the limitation of circular symmetry. Combined with the optimized surface discretization method such as the isoenergetic grid, we have shown that the generalized phase space method can be applied in the surface scattering simulation of optical systems with arbitrary geometry. According to the geometries of the optical systems that need to be simulated and the target values of the simulation, we have proposed three different implementations of the phase space method in surface scattering simulation, and the applicability of these implementations have been illustrated by two examples. Meanwhile, the accuracy and efficiency of these implementations have been discussed in detail, as well as the disadvantages of the method such as the lack of robustness. Based on the discussion, an automatic determination of the method for different simulation scenarios is possible.

At last, we have demonstrated the application of the Wigner function-based method to propagate partially coherent beams in an EUV Schwarzschild objective and a retrofocus lens with both the MSF and HSF surface error on the two elements. These examples show that the Wigner function-based method is able to combine surface scattering from all frequency components of the PSD with diffraction and interference in a unified model. However, the Wigner function-based simulation is currently limited to 2D, because the extension to the 3D space will result in 4D Wigner function, which consume large memory and is extremely computationally intensive.

In general, the work in this dissertation establishes a new physical model that solves the problem of traditional straylight simulation methods in efficiency, SNR and physical propagation. With the help of the new method, straylight analysis, which is one of the most time-consuming part in optical design process, can be significantly accelerated. Therefore, the optical designers can have timely feedbacks on the straylight performance of the optical system at an early stage of optical design, which makes it possible to integrate the straylight performance throughout the optimization of optical systems. Additionally, the Wigner function-based physical propagation method allows the calculation of the PSFs of as-built optical systems, which is essential for the surface tolerancing of optical systems that involve partially coherent light.

Appendix A: Isoenergetic sampling of incident rays by the Fibonacci grid

The Fibonacci grid in the pupil guarantees quasi-isoenergetic sampling of the illumination ray bundle. Depending on whether the object space is finite or afocal, angular or spatial sampling by the Fibonacci grid should be applied.

If the object space is finite, the rays should be uniformly distributed on the spherical cap defined by the incident ray cone. In this case, the deflection angle θ and azimuthal angle φ of the rays sampled on the Fibonacci grid are calculated as [42]:

$$\begin{aligned}\varphi_i &= i \cdot \left[\pi \cdot (3 - \sqrt{5}) \right] \\ \theta_i &= \arccos \left[1 - \frac{i}{N} \left(1 - \sqrt{1 - NA^2} \right) \right],\end{aligned}\tag{A.1}$$

where i is the index of the ray, N is the total number of sampled rays and NA is the object space numerical aperture of the optical system.

In case of afocal object space, the rays should be uniformly distributed on the entrance pupil, and the spatial coordinates of the rays are calculated in the polar coordinate system as follows:

$$\begin{aligned}r_i &= \frac{D}{2} \sqrt{\frac{i}{N}} \\ \varphi_i &= i \cdot \pi \cdot (3 - \sqrt{5}),\end{aligned}\tag{A.2}$$

where i is the index of the ray, N is the total number of sampled rays and D is the entrance pupil diameter. Fig. shows a Fibonacci grid with 1000 sampling points calculated with Eq. (A.2).

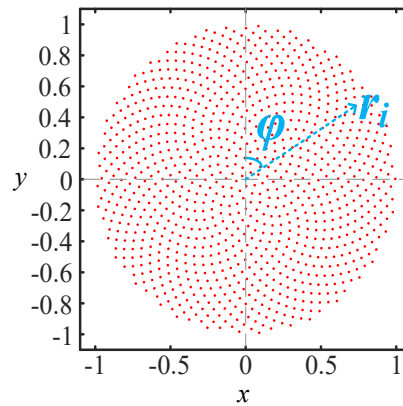


Fig. A-1. A Fibonacci grid with 1000 sampling points.

Furthermore, it should be noted that the Fibonacci sampling is quasi-isoenergetic, meaning that the sampling is isoenergetic only when the Fibonacci grid is denser than the

grid on which the optical surface is discretised. In case the density of the Fibonacci grid is comparable to that of the surface discretization, isoenergetic sampling is not obtained due to the Moore effect. In our simulations we have fixed the density of the Fibonacci grid such that there are at least 50 rays in each subarea of the optical surface. Once this condition is fulfilled, the sampling of the entrance pupil is considered to be isoenergetic and apodizations can be modelled by assigning weightings to the incident rays according to their pupil coordinates.

References

1. W. Singer, M. Totzeck, and H. Gross, *Handbook of Optical Systems, Volume 2, Physical Image Formation*, Vol. 2: Wiley-VCH, 2005.
2. X. Lu, O. Rodenko, Y. Zhang, and H. Gross, *Efficient simulation of autofluorescence effects in microscope lenses*, *Applied optics*, **58**(13): p. 3589-3596, 2019.
3. J.C. Stover, *Optical scattering: measurement and analysis*: Society of Photo-Optical Instrumentation Engineers, 2012.
4. J.E. Harvey, *Understanding Surface Scatter Phenomena: a linear systems formulation*: SPIE Press, 2019.
5. G.L. Peterson. *Analytic expressions for in-field scattered light distributions*. in *Optical Modeling and Performance Predictions*. International Society for Optics and Photonics, 2004
6. K. Liang, G. Forbes, and M.A. Alonso, *Validity of the perturbation model for the propagation of MSF structures in 3D*, *Optics Express*, **28**(14): p. 20277-20295, 2020.
7. J.H. Seinfeld and S.N. Pandis, *Atmospheric chemistry and physics: from air pollution to climate change*: John Wiley & Sons, 2016.
8. G. Mie, *Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen*, *Annalen der physik*, **330**(3): p. 377-445, 1908.
9. J. Pawley, *Handbook of biological confocal microscopy*, Vol. 236: Springer Science & Business Media, 2006.
10. SCHOTT, *TIE-36: Fluorescence of optical glass*, 2010.
11. A. Krywonos, *Predicting surface scatter using a linear systems formulation of non-paraxial scalar diffraction*, Ph. D. Dissertation, University of Central Florida, 2006.
12. M.G. Dittman. *K-correlation power spectral density and surface scatter model*. in *Optical Systems Degradation, Contamination, and Stray Light: Effects, Measurements, and Control II*. International Society for Optics and Photonics, 2006
13. J.E. Harvey. *Light-scattering characteristics of optical surfaces*. in *Stray Light Problems in Optical Systems*. International Society for Optics and Photonics, 1977
14. A. Papoulis, *Probability and statistics*: Prentice-Hall, Inc., 1990.
15. R. Irvin. *Applying Importance Sampling Specifications in Script*. Available from: <https://photonengr.com/applying-importance-sampling-specifications-in-script>. 2010. Last
16. A.S. Ostrovsky, *Coherent-mode representations in optics*, Vol. 164: SPIE Press, 2006.
17. M. Testorf. *Phase-space optics applied to the design of Talbot array illuminators*. in *Diffraction Optics and Micro-optics*. Optical Society of America, 2004
18. J.W. Goodman, *Statistical optics*: John Wiley & Sons, 2015.
19. E. Wolf, *New theory of partial coherence in the space-frequency domain. Part I: spectra and cross spectra of steady-state sources*, *JOSA*, **72**(3): p. 343-351, 1982.
20. D. Dragoman, *Phase-space interferences as the source of negative values of the Wigner distribution function*, *JOSA A*, **17**(12): p. 2481-2485, 2000.

21. R. Shi, W. Kleinschmidt, and T. Bocher, *Microscope objective*, U.S. Patent 7907348, Carl Zeiss MicroImaging GmbH, 2011.
22. Y. Zhang, *Systematic design of microscope objectives*, Ph. D. Dissertation, Friedrich-Schiller-Universität Jena, 2019.
23. H. Konishi, Y. Fujimoto, K. Kusaka, and T. Kasahara, *Immersion type microscope objective*, U.S. Patent 7486445 B2, Olympus Corp, 2009.
24. R. Wartmann and J. Sprenger, *Liquid immersion microscope objective*, U.S. Patent 20060018030, Carl Zeiss Microscopy GmbH, 2006.
25. I. Fahlbusch and R. Shi, *Immersion microscope objective*, U.S. Patent 20060087745A1, Carl Zeiss Microscopy GmbH, 2006.
26. K. Watanabe, *Microscope objective lens*, Japanese Patent 2005-189732, Olympus Corp, 2005.
27. T. Otaki, T. Nihoshi, M. Sato, H. Kaizu, et al., *Microscope system*, U.S. Patent 6128128A, Nikon Corp, 2005.
28. R. Jedamzik, F. Elsmann, A. Engel, U. Petzold, et al. *Introducing the quantum efficiency of fluorescence of SCHOTT optical glasses*. in *Current Developments in Lens Design and Optical Engineering XVIII*. International Society for Optics and Photonics, 2017
29. S. Reichel, R. Biertümpfel, and A. Engel. *Characterization and measurement results of fluorescence in absorption optical filter glass*. in *Optical Systems Design 2015: Optical Design and Engineering VI*. International Society for Optics and Photonics, 2015
30. Y. Zhang and H. Gross, *Systematic design of microscope objectives. Part II: Lens modules and design principles*, *Advanced Optical Technologies*, **8**(5): p. 349-384, 2019.
31. SCHOTT. *D 263® T Thin Glass*. Available from: https://www.schott.com/d/advanced_optics/ea165ca6-b4f8-49ee-b989-072ddb2d2e4d/1.4/schott-d-263-t-eco-thin-glass-may-2013-eng.pdf. May 2013. Last accessed 2020-12-14
32. SCHOTT. *BOROFLOAT® 33 – Optical Properties*. Available from: https://www.us.schott.com/d/borofloat/1ca1519d-b5b3-469d-b3eb-9b80235b1783/1.4/borofloat33_opt_usa_web.pdf. 2014. Last accessed 2020-12-14
33. M. Zhong, *Propagation of partially coherent light in optical systems*, Doctoral Thesis, Friedrich-Schiller University of Jena, 2016.
34. U. Zastra, C. Rödel, M. Nakatsutsumi, T. Feigl, et al., *A sensitive EUV Schwarzschild microscope for plasma studies with sub-micrometer resolution*, *Review of Scientific Instruments*, **89**(2): p. 023703, 2018.
35. A.E. Siegman, *Lasers university science books*, Mill Valley, CA, **37**(208): p. 169, 1986.
36. H. Lajunen, J. Tervo, J. Turunen, T. Vallius, et al., *Simulation of light propagation by local spherical interface approximation*, *Applied optics*, **42**(34): p. 6804-6810, 2003.
37. S. Cho, J. Petrucci, and M. Alonso, *Wigner functions for paraxial and nonparaxial fields*, *Journal of Modern Optics*, **56**(17): p. 1843-1852, 2009.

38. Y. Zhang and B. Lü, *Propagation of the Wigner distribution function for partially coherent nonparaxial beams*, Optics letters, **29**(23): p. 2710-2712, 2004.
39. L. AW, J. Ojeda-Castañeda, and N. Streibl, *The influence of wave aberrations on the Wigner distribution*, Optica Applicata, **13**(4)1983.
40. H. Gross, *Numerical propagation of partially coherent laser beams through optical systems*, Optics & Laser Technology, **29**(5): p. 257-260, 1997.
41. D. Dragoman, *Wigner distribution function applied to third-order aberrations*, Applied optics, **35**(1): p. 161-168, 1996.
42. C. Drost. *The golden spiral method*. Available from: <https://stackoverflow.com/questions/9600801>. May 2017. Last accessed 2020-07-15

List of figures

- Fig. 2-1. Polar diagrams of the scattering phase function of a spherical scatterer with a refractive index of $n = 1.33$ calculated by the Mie solution, $q = 2\pi/\lambda$ 5
- Fig. 2-2. Relative fluorescence coefficients of some Schott glasses normalized on the value of SF1. The excitation is at $\lambda = 365$ nm and the signal is spectrally integrated [10]. 6
- Fig. 2-3. Log-log scaled power spectral density (PSD) of optical surfaces manufactured by surface grinding. The figure error is the low spatial frequency (LSF) part while the micro roughness is the high spatial frequency (HSF) part of the PSD. 7
- Fig. 2-4. Illustration of the surface height variation of three types of surface errors. The surface variations are modelled as Gaussian processes with different correlation lengths and RMS surface height. Only part of the full aperture is shown for clear illustration. 7
- Fig. 2-5. Schematic sketch of an imaging optical system with scattering surfaces, $P_i(\zeta, \eta)$ is the intrinsic pupil function that includes all the geometrical aberrations, apodizations and edge diffractions, while $P_s(\zeta, \eta)$ is the additional pupil function that encapsules the phase variation at the exit pupil induced by the residual surface errors. 8
- Fig. 2-6. 1-dimentional PSFs of optical systems that contains the surface errors of Fig. 2-4. a) shows the PSFs near the central peak. b) shows the PSFs in a larger spatial range. c) shows the side lobes of the PSFs far from the central peak. d) shows the encircled energy of the PSFs. 9
- Fig. 2-7. Modulation transfer function (MTF) of the optical systems with residual surface errors. The solid lines correspond to the actual MTFs of the surface error 1-3 shown in Fig. 2-4. The dashed lines indicate to the MTFs predicted by the Harvey model. 10
- Fig. 2-8. Observation of the diffracted field by a diffracting aperture [13]. 12
- Fig. 2-9. Phase modulation by surface height variations of a scattering surface for oblique incidence and scattering angle. The green segments correspond to the optical path difference (OPD) induced by the surface height h 16
- Fig. 2-10. Simulation of volume scattering in a microscope lens (a) and surface scattering in a Kirkpatrick-Baez telescope (b) by Monte Carlo raytracing in OpticStudio. As an illustration, 10 rays are launched into the optical systems from the on-axis object in each system. 18
- Fig. 2-11. Monte Carlo raytracing without importance sampling (a) and with importance sampling (b) for a TMA system. The target object of importance sampling is the detector. M3 is a scattering surface with a Gaussian BRDF. 20
- Fig. 2-12. Wigner functions of a top-hat Gaussian-Schell beam after a positive lens and free-space propagation. 22
- Fig. 3-1. Illustration of the generation of straylight from a scattering surface. a) shows the layout of the setup. b) and c) show the phase space distribution of the illumination and scattered light. d) shows the acceptance of the

detector in the phase space. e) shows the overlapping between the generation and acceptance of scattered light.....	24
Fig. 3-2. Sketch of a simplified model of a microscope lens.....	26
Fig. 3-3. Coupling of the illumination and acceptance in a subarea of a reflective surface. The red ray is traced from the object space while the blue rays are traced from the four corners of the detector. The pentahedron cone formed by the four blue rays defines the solid acceptance angle.....	27
Fig. 3-4. Generation of straylight in a subarea of an optical surface. O is an on-axis object point and O' is on the intermediate image plane of the detector (details to follow in Sec. 5.1). The red mesh on the optical surface indicates the isoenergetic polar grid used to discretize the surface, and the color code indicates the relative irradiance distribution on the optical surface due to Gaussian apodization. The color code on the intermediate image plane of the detector indicates the contribution of the surface subarea to the irradiance distribution on the detector.	29
Fig. 3-5. Intermediate images of the detector seen by different subareas in the presence of spherical aberrations. The green mesh is the intermediate image seen by the off-axis point A , while the red curve is that seen by the on-axis point O	30
Fig. 3-6. The field coordinate y and pupil coordinates (r_p, φ) of an arbitrary subarea-pixel combination. The red and blue areas indicate the pixel and surface subarea. For simplicity, the chief ray is assumed to intersect with the optical surface at the optical axis, while in general cases, the pupil coordinates should be referred to the actual intersection point of the chief ray with the surface. For each pixel, the transverse aberration is evaluated in the local coordinate system in which the y axis is parallel to the field vector as indicated by the red dashed line.....	32
Fig. 3-7. Distorted detector grid due to aberrations seen by a point on the optical surface. The red dots indicate the locations of pixel centers in the distorted intermediate image while the blue arrows indicate the shift of pixel centers due to aberrations. The wave aberrations are represented by Seidel coefficients. a) shows spherical aberration with $S_I = -0.1$ mm; b,c) show coma with $S_{II} = -0.1$ mm; d) shows astigmatism with $S_{III} = -0.1$ mm; e) shows Petzval curvature with $S_{IV} = -0.2$ mm; f) shows distortion with $S_V = 0.2$ mm. c) corresponds to a sagittal point on the optical surface while the rest correspond to tangential points. The size of the paraxial intermediate image is $4.9 \text{ mm} \times 4.9 \text{ mm}$	33
Fig. 3-8. Impact of aberrations on the intermediate image of the detector. The arrows indicate the intermediate images seen by different points on the optical surface.	34
Fig. 3-9. Schematic illustration of the inverse raytracing and hybrid methods. The dashed grid in the detector intermediate image shows to the discretization of the detector intermediate image in the hybrid method.	34
Fig. 3-10. Flowchart of the three methods to implement the phase space model.....	36
Fig. 3-11. Schematic sketch of the simulation of volume scattering by slicing the optical system along the optical axis. The blue ray cone is the	

- illumination ray cone, and the red ray cone is the acceptance ray cone of the detector for the central subarea of the z-slice. 37
- Fig. 3-12. Schematic sketch of a retrofocus lens in a reversed order. The blue dotted arrow shows the intermediate image of the detector seen by L1. Points O and B are an on-axis point and an off-axis point on L1, and the orange and pink arrows indicate the portions of the intermediate image seen by these two points. 38
- Fig. 3-13. Propagation of a Gaussian Schell beam through a scattering surface at $z = 3 \text{ mm}$ with HSF surface errors. The first three figures show the Wigner function of the beam before the scattering surface (a), after the scattering surface (b), and after 30 mm of free space propagation (c). d) shows the irradiance distribution of the beam. 44
- Fig. 3-14. An example of surface structure of MSF surface error. Displacement of the tool tip due to vibrations in thrust and feed directions follows Gaussian distributions, and phase delay $\Delta\phi$ is calculated for the wavelength of $1 \mu\text{m}$ 46
- Fig. 3-15. Propagation of a Gaussian Schell beam through a scattering surface with MSF structures and a single lens. a) shows the irradiance distribution on different z-planes for the scattering surface described in Fig. 3-14, the single lens having a focal length of $f = 40 \text{ mm}$. b) shows the x-cross-section of the irradiance distribution at the beam focus for four MSF structures with different periods and PV values. 46
- Fig. 3-16. Wigner functions of the Gaussian Schell beam propagated through the optical system described in Fig. 3-15(a) for an MSF structure with $T = 40 \mu\text{m}$ and $h_{pv} = 0.5 \mu\text{m}$ 47
- Fig. 4-1. Schematic setup of the model with excitation light coming from the object plane, total fluorescent light created in the lens and detected fluorescence light reaching one detector pixel. 49
- Fig. 4-2. Sketch of the spatial and angular overlap of the excitation (yellow), fluorescence (blue) and detected (red) light in the phase space in one dimension. The efficiency of transfer is given by the common area between the detection and fluorescent light, which is seen in red. 50
- Fig. 4-3. Layout (a) and simulation result (b) of a plastic projection lens. The simulation results are calculated geometrically with the phase space and Monte Carlo raytracing without considering the autofluorescence coefficient of the plastic material. 52
- Fig. 4-4. Layout and simulation result of a $100\times/\text{NA } 1.25$ microscope lens [21]. The simulation results are calculated by the phase space and brute-force method without considering the autofluorescence coefficient of the glass, the orange curve corresponds to the relative error of the phase space method. Color-coding of the layout is not related to the autofluorescence coefficient of glass. 53
- Fig. 4-5. Excitation area (a), acceptance area (b) and acceptance angle (c) calculated by paraxial ray tracing and real ray tracing for the first element. 54
- Fig. 4-6. Intensity distribution in the first element of the $10\times/\text{NA } 1.25$ microscope lens at $z = 0.1 \text{ mm}$ (a) and $z = 0.7 \text{ mm}$ (b). 55

- Fig. 4-7. Real phase space acceptance area and ideal acceptance area without aberration and vignetting. 57
- Fig. 4-8. Dependence of the fluorescent intensity on image height. 57
- Fig. 4-9. Comparison of phase-space-based model with Zemax for the example system in Fig. 4-4(a). The calculation with the paraxial model was performed for full NA. 58
- Fig. 4-10. Simulated relative geometrical autofluorescence intensity vs. system etendue, the maximum value has been normalized to 1, the magnifications of the microscope lenses are indicated by the color code. The etendue is defined as $y^2 \cdot NA^2$, where y is the object height and NA is the object space numerical aperture. 59
- Fig. 4-11. Geometrical autofluorescence intensity vs. system etendue for high NA, high magnification (a) and low NA, low magnification (b) lenses. The color code represents $1/wd^2$, where w is the working distance of the microscope lens. The etendue is defined as $y^2 \cdot NA^2$, where y is the object height and NA is the object space numerical aperture. 60
- Fig. 4-12. Autofluorescence coefficients of optical glasses with an excitation wavelength of 365nm. a) shows all glasses with available data while b) shows the lead-free ('N-type') glass, the color code corresponds to the logarithmically-scaled autofluorescence coefficients, which are normalized by the maximum value in each plot. The number of glass types in b) is less than that in a) since the glasses shown in b) is a subset of that in a). 62
- Fig. 4-13. Dependence of autofluorescence coefficients on the anomalous partial dispersion ($\Delta P_{g,F}$) for lead-free ('N-type') glasses. 63
- Fig. 4-14. Relative frequency of use of glasses in the glass map for the front (a), middle (b) and rear (c) groups. Gaussian smoothing is applied to each figure and the peak values are normalized to 1. 63
- Fig. 4-15. Layouts and autofluorescence contribution from the elements of lens (i) and (ii). The light source is assumed to be an on-axis point source located at the object plane, emitting excitation light at 365nm isotropically in all directions. The power of the light source is 1 W. The fluorescent light is measured on the detector plane, the size of the detector is 11mm \times 11mm, the autofluorescence contribution from an element is characterized as the irradiance of fluorescent light on the detector. The color coding shows the logarithmically scaled relative autofluorescence coefficients, and the maximum value is normalized to 1. The results are based on a tube lens with a focal length of 200mm. 64
- Fig. 4-16. Autofluorescence contribution from the first element vs. the index-weighted working distance. 66
- Fig. 4-17. Layout of lens (ii) and the z-sectioning of its first element in the phase-space-based simulation. The intermediate image in the first element is marked in order to indicate its distance to each slice. b) z-dependence of autofluorescence contribution from the first element. The horizontal axis is the z-location of a slice relative to the intermediate image plane, the red curve corresponds to the relative autofluorescence contribution from each slice, while the green curve corresponds to the accumulated

autofluorescence contribution of all slices whose distance to the intermediate image is smaller than z	67
Fig. 4-18. Autofluorescence contribution from different types of cover glass in comparison with the first element of lens (ii), the thickness of the cover glass is 0.17mm.	68
Fig. 4-19. a) Layout of lens (i) and z-sectioning of its rear element in the phase-space-based simulation. The intermediate image in the 9 th element is marked in order to indicate its distance to each slice. b) z-dependence of autofluorescence contribution from element 9, the horizontal axis is the z-location of a slice relative to the intermediate image plane, the red curve corresponds to the relative autofluorescence contribution from each slice, while the green curve corresponds to the accumulated autofluorescence contribution of all slices whose z-coordinate is larger than z , with the intermediate image plane being the origin.....	69
Fig. 4-20. Glass substitution based on the glass map.	70
Fig. 4-21. Layouts and autofluorescence contribution of lenses (i) and (ii) after glass substitution. The light source and normalization are the same as described in Fig. 4-15.	71
Fig. 4-22. Splitting of the rear group of lens (i). a) shows the original structure of the rear group, b) shows the structure with a planar cementing surface and c) shows the structure with a curved cementing surface. $w220$ and $w222$ are the Seidel coefficients for Petzval curvature and astigmatism in waves, here it represents the overall Petzval curvature and astigmatism contribution of the rear group. From the values of the Seidel coefficients, it is obvious that the aberration contributions from the initial (a) and final (c) systems are very similar, while the Petzval curvature contribution of (b) greatly differs from that of the initial system.	72
Fig. 4-23. Splitting of the front group of lens (ii), the two layers in the front of element 1 are the cover glass and immersion medium. a) shows the original structure of the front group, b) shows the structure with a planar cementing surface in element 1 and thick meniscus lens in element 2, c) shows the structure with a curved cementing surface in element 1 and thin shell lenses in element 2. $w040$ is the Seidel coefficient for spherical aberration in waves, here it represents the overall spherical aberration contribution of the front group.	72
Fig. 4-24. Layouts and autofluorescence contribution of lenses (i) and (ii) after glass substitution and lens splitting. The light source and normalization are the same as described in Fig. 4-15.....	73
Fig. 5-1. Layout of the Ritchey–Chrétien telescope.....	75
Fig. 5-2. Irradiance distribution on M1 and M2 for an incident beam of 1W. The red mesh indicates the isoenergetic grid on which M1 and M2 are discretized.....	76
Fig. 5-3. Coupling coefficients between two tangential subareas on M1 and M2 with the intermediate images of the pixels. The x and y coordinates correspond to the positions of the pixel centers in the intermediate images.....	77

- Fig. 5-4. Point source transmittance (PST) function of M1 and M2 for an incident beam of 1 W. The x and y coordinates correspond to the pixel centers on the 100×100 detector. The color code indicates the scattered irradiance on the detector, the units are W/mm^2 78
- Fig. 5-5. x cross-sections of the PSTs calculated by the phase space model and the Monte Carlo method. The relative RMS deviation is normalized to the peak values in each figure. 78
- Fig. 5-6. Layout of the Kirkpatrick-Baez telescope. 79
- Fig. 5-7. Irradiance and solid acceptance angle distribution on M1 and M2 in local coordinate systems. 80
- Fig. 5-8. Total scattered flux on the detector contributed by M1 and M2. The results are calculated by the phase space model and Monte Carlo method. 81
- Fig. 5-9. PSTs of the KBT for an incident beam from the on-axis field with a power of 1 W. The detector is a 100×100 square pixel array with a width of 2.6 mm. 81
- Fig. 5-10. PSTs of the KBT calculated by the Monte Carlo method. The dark holes in the centers of the first two figures result from the element that blocks the specular rays. The bottom figures show the comparison of the x and y cross-sections of the PSTs calculated by the Monte Carlo method and the phase space model. The Relative RMS deviation is normalized to the peak values in each figure. 82
- Fig. 5-11. a) Layout of a single lens imaging system. b) RMS relative error of the simulation results vs. the number of rings in the isoenergetic polar sampling grid. c) Log-log plot of the RMS relative error of the simulation results vs. runtime. The errors are determined relative to the results calculated by the intermediate image method with 20 rings. 84
- Fig. 5-12. A simplified two-element retrofocus lens. The incident Gaussian-Schell beam is truncated by the front element and the stop. The distances between the stop and the two lenses are 15mm. 86
- Fig. 5-13. Wigner distributions of the beam in the retrofocus lens. 87
- Fig. 5-14. a) Irradiance distribution of the beam in the retrofocus lens. b) the x -cross-section of the irradiance distribution at the beam focus. 87
- Fig. 5-15. An EUV Schwarzschild objective with an object space numerical aperture (NA) of 0.19 and an image space NA of 0.004. Only part of the ray bundle in the image space are shown due to visualization purposes. The diameters of M1 and M2 are 50 mm and 10.6 mm respectively. 88
- Fig. 5-16. a) 1-dimensional Irradiance distribution of the beam from the entrance window to M1 and M2. b) 1-dimensional Irradiance distribution from M2 to the image plane. c) 1-dimensional irradiance distribution in the image plane. 89
- Fig. 5-17. Wigner distributions of the beam in the Schwarzschild objective after the entrance window (a), M1 (b), and M2 (c). The parabolic wavefronts of the beams have been removed. 90

List of tables

Table 5-1. Surface data of the Kirkpatrick-Baez telescope.....	79
Table 5-2. Comparison between the three methods of the phase space model.....	83

List of abbreviations

SNR	Signal to noise ratio
BSDF	Bi-directional scattering distribution function
BRDF	Bi-directional reflection distribution function
MSF	Mid-spatial frequency
HSF	High spatial frequency
PSD	Power spectral density
PSF	Point spread function
MTF	Modulation transfer function
HS	Harvey-Shack theory
ASF	Angular spread function
OHS	Original Harvey-Shack theory
GHS	Generalized Harvey-Shack theory
NA	Numerical aperture
TIS	Total integrated scattering
TMA	Three-mirror astigmat
FOV	Field of view
TEA	Thin element approximation
LSIA	local spherical interface approximation
PV	Peak-valley
EUV	Extreme ultraviolet
RCT	Ritchey–Chrétien telescope
KBT	Kirkpatrick-Baez telescope

PST	Point source transmittance
ACV	Autocovariance

List of symbols

δ_s	Scattering cross-section
μ_s	Scattering coefficient
l_c	Correlation length
σ_s	Standard deviation of surface error
$A(x,y,z)$	Amplitude of the light field
$H(u,v)$	Transfer function
$E \{ \}$	Ensemble average
λ	Wavelength of light
$P_s(\xi,\eta)$	Complex pupil function induced by residual surface error
α	x -directional cosine
β	y -directional cosine
ϕ	Radiant flux
I	Irradiance intensity
$W_R(x',y')$	Surface height variation
\otimes	Convolution
$\Phi(\omega_1, \omega_2)$	Joint characteristic function
$C_w(u, \hat{v})$	Autocovariance of the surface height variation
f	Spatial frequency
$\Gamma(x, \Delta x)$	Autocorrelation function
$t(x)$	Amplitude modulation of the thin element
$W(x,u)$	Wigner distribution function
$d\Omega$	Solid acceptance angle

n	Refractive index
x_p, y_p	Pupil coordinates
$\mathbf{J}(P_1, P_2)$	Mutual intensity function defined by time-average
$\bar{\mathbf{J}}(P_1, P_2)$	Mutual intensity function defined by ensemble-average
T	Period of MSF structure
h_{pv}	PV value of MSF structure
σ_t	Standard deviations of the tool tip vibration in the thrust direction
σ_f	Standard deviations of the tool tip vibration in the feed direction
η_{flu0}	Autofluorescence coefficient
g_{1d}	1D geometrical coupling coefficient of straylight
g_{2d}	2D geometrical coupling coefficient of straylight
$\mathbf{u}(P, t)$	Electro-magnetic field at point P , time t
u, v	Angular directions
φ_p	Phase modulation by a scattering surface

Acknowledgement

I have imagined this moment many times, because every time I received help and support during the last four years, it had always brought this moment to my mind. The completion of this work will never be possible without the help from the many persons, to whom I owe a great debt of gratitude.

I would like to thank my supervisor, Prof. Herbert Gross. Since our acquaintance in 2014, he has always been inspiring me with his abundant knowledge and enthusiasm towards profession. His patience and generosity allowed me to explore freely in the academic field, and he is always there to help and guide when I feel lost. Through the six years, he showed me how to be a good teacher, a good supervisor, a good scientist, a good colleague and a good friend. I believe that what I learnt from him will always benefit me in the future.

I also feel grateful to be around so many great group members. Their willingness to help and unique sense of humor have created a wonderful atmosphere in the group. Our insightful discussions have sparked a large fraction of my inspirations and ideas. I want to thank them for giving me so much support, and at the same time giving me the chance to help others. It is extremely fortunate for me to be accompanied by these lovely people through this journey. I, and my stomach will never forget about our movie nights, BBQ and stollen parties, as well as our countless afternoon cake times, these good memories will be treasured by me forever.

Last but not least, I want to thank my family, especially my wife and my parents, for always being there for me and all the selfless support and trust.

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbständig, ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen:

- Herbert Gross, Betreuer.

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die geltende Promotionsordnung der Physikalisch-Astronomischen Fakultät ist mir bekannt.

Ich versichere ehrenwörtlich, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Publications

Peer-reviewed Journals

X. Lu, O. Rodenko, Y. Zhang and H. Gross,
“Efficient simulation of autofluorescence effects in microscope lenses”,
Applied Optics 58, 3589-3596 (2019).

X. Lu, Y. Zhang, and H. Gross,
“General analysis and optimization strategy to suppress autofluorescence in
microscope lenses”,
Applied Optics 58, 7404-7415 (2019).

X. Lu and H. Gross,
“Efficient simulation of surface scattering in symmetry-free optical systems”
Optics Express 28, 39368-39386 (2020).

X. Lu and H. Gross,
“Wigner function-based modeling and propagation of partially coherent light in optical
systems with scattering surfaces”
Optics Express 29, 14985-15000 (2021).

Conference proceedings

X. Lu, R. Heintzmann and U. Leischner,
“High-resolution light-sheet microscopy: a simulation of an optical illumination system
for oil immersion”
Proceedings of SPIE Vol. 9626, p.96262I (2015).

X. Lu and H. Gross,

“Simulation of autofluorescence in microscope lenses”,

SAOT International Conference on Advanced Optical Technologies (2019).

B. Messerschmidt, G. Matz , H. Gross, X. Lu, et al.

“Dual modality multiphoton-OCT flexible endomicroscope with an integrated electromagnetic z-actuator for optical field-of-view switching and a piezo-fiber-scanner for image acquisition”.

Proceedings of SPIE Vol. 11214, p. 112140X (2020).

Talks

X. Lu, R. Heintzmann and U. Leischner,

“High-resolution light-sheet microscopy: a simulation of an optical illumination system for oil immersion”

Optical Systems Design 2015: Optical Design and Engineering VI, Jena (2015).

X. Lu and H. Gross,

“Simulation of autofluorescence in microscope lenses”,

SAOT International Conference on Advanced Optical Technologies, Erlangen (2019).