



**FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA**

**DEVELOPMENT OF
FREEFORM 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. **Danyun Cai**

Born in Anqing, Anhui Province, P. R. China on 11.03.1992

Gutachter:

1. Prof. Dr. Herbert Gross, Friedrich-Schiller-Universität Jena
2. Prof. Dr. Fabian Dürr, Vrije Universiteit Brussel
3. Prof. Dr. Alois Herkommer, Universität Stuttgart

Date of the disputation: 03.11.2022

Zusammenfassung

Mit der Verbesserung der optischen Fertigungsmöglichkeiten werden heute Freiformflächen in verschiedenen optischen Systemen eingesetzt, um die Bildqualität zu verbessern, das Sichtfeld zu vergrößern, und die Systemgröße zu verringern. Eine der wesentlichen Eigenschaften der Freiformflächen ist, dass ihr asymmetrisches und lokal abweichendes Oberflächenprofil die Symmetrie eines optischen Systems bricht und damit neuartige Probleme aufwirft, die bei rotationssymmetrischen optischen Systemen noch nie berücksichtigt wurden. Diese Arbeit konzentriert sich auf die Entwicklung eines Optimierungsalgorithmus, der automatisch die Mittenabschattung in einem nicht-rotationssymmetrischen reflektierenden optischen System beseitigt, sowie auf die Definition und Berechnung der verallgemeinerten chromatischen Aberrationen in einem nicht-rotationssymmetrischen refraktiven optischen System. Darüber hinaus wird ein umfassendes Modell für die Tolerierung von Freiformflächen vorgestellt.

Bei der Optimierung eines dreidimensionalen (3D) reflektierenden optischen Systems durch Kippen der Spiegel können die Spiegel den Strahlengang teilweise blockieren und infolgedessen die Bildhelligkeit und den Kontrast verringern. Um den Grad der Abschattung zu berücksichtigen, wird eine Fehlerfunktion vorgeschlagen, die alle geometrisch auftretenden Fälle in reflektierenden 3D-Systemen mathematisch beschreibt. Die Fehlerfunktion kann mit der Bildqualitätsfehlerfunktion kombiniert werden, um gleichzeitig die Aberrationen zu kontrollieren und die Vignettierung während des Optimierungsprozesses zu beseitigen.

Um die verallgemeinerten chromatischen Aberrationen in refraktiven 3D-Systemen zu analysieren, werden die Referenzachse und die Referenzebene geklärt, um die genaue Definition der verallgemeinerten chromatischen Aberrationen herauszufinden. Es werden sowohl strahlenbasierte als auch wellenfrontbasierte Methoden vorgeschlagen, um die verallgemeinerten chromatischen Aberrationen flächenaufgelöst in einem System zu berechnen. Darüber hinaus wird der Einfluss der Pupillenaberration erörtert, um die Berechnungsgenauigkeit zu verbessern. Die flächenaufgelösten verallgemeinerten chromatischen Aberrationen werden in einer zweidimensionalen feldabhängigen Form graphisch dargestellt.

Der Fertigungsfehler von Freiformflächen kann durch Fourier-Transformation in den Frequenzbereich übertragen werden. Die PSD-Kurve steht in direktem Zusammenhang mit

Zusammenfassung

der RMS-Rauheit. Die Autokorrelationsfunktion (ACF) des Phasenmusters wird in verschiedenen Frequenzbereichen berechnet. Durch die Charakterisierung der Breite der ACF kann die Grenzfrequenz und den Übergang zwischen den deterministischen und statistischen Fehlern gefunden werden. Es wird ein umfassendes Modell vorgeschlagen, das verschiedene Arten von Oberflächenfehlern darstellt, um eine synthetische Freiformfläche zu konstruieren. Durch die Durchführung der Monte-Carlo-Simulation kann damit die Tolerierung des Freiformsystems realisiert werden.

Abstract

Nowadays, with the improvement of optical manufacturing capability, freeform surfaces are widely used in different kinds of optical systems to enhance the image quality, enlarge the field of view, reduce the system size, and so on. One of the essential properties of the freeform surface is that its asymmetric and locally variant surface profile breaks the symmetry of an optical system, thus provoking unique issues that have never been considered in rotationally symmetric optical systems. This thesis focuses on developing an optimization algorithm that automatically eliminates the obscuration in the non-rotationally symmetric reflective optical system, as well as defining and computing the generalized chromatic aberrations in the non-rotationally symmetric refractive optical system. Furthermore, a comprehensive model for the tolerancing of freeform surface is put forward.

When optimizing a three-dimensional (3D) reflective optical system by tilting the mirrors, the mirrors can block the ray path and, in consequence, reduce the image brightness and contrast. To take the degree of obscuration into consideration, an error function that mathematically describes all the obscuration cases in 3D reflective systems is proposed. The obscuration error function can be combined with the image quality error function to simultaneously control the aberrations and eliminate the obscuration during the optimization process.

In order to analyze the generalized chromatic aberrations in 3D refractive systems, the reference axis and reference plane are clarified to figure out the precise definition of the generalized chromatic aberrations. Both ray-based and wavefront-based methods are proposed to calculate the generalized chromatic aberrations surface-by-surface. In addition, the influence of pupil aberration is discussed to improve calculation accuracy. The surface-resolved generalized chromatic aberrations are presented in the full-field display.

The manufacturing error of the freeform surface can be transferred into the frequency domain by Fourier transform. The power spectrum density curve directly relates to the root-mean-square roughness. The autocorrelation function (ACF) of the phase pattern is computed in different frequency ranges. By characterizing the width of ACF, the boundary frequency between the deterministic and statistic errors can be found. A comprehensive model representing different types of surface errors is proposed to construct a synthetic freeform surface. By performing the Monte-Carlo simulation, tolerancing of the freeform system can be realized.

This page is left blank intentionally.

Contents

| | |
|--|-----|
| Zusammenfassung | i |
| Abstract | iii |
| Contents | 1 |
| 1 Motivation | 3 |
| 2 Theoretical Introduction | 7 |
| 2.1 Optical surfaces | 7 |
| 2.1.1 Basic shapes | 7 |
| 2.1.2 Aspheres | 8 |
| 2.1.3 Freeforms | 10 |
| 2.2 Classification of optical systems | 10 |
| 2.3 Single surface imaging equations | 14 |
| 2.4 Aberration theories | 17 |
| 2.4.1 Aberration theories in rotationally symmetric optical systems | 18 |
| 2.4.2 Aberration theories in non-rotationally symmetric optical systems | 20 |
| 2.5 Obscuration in off-axis mirror systems | 21 |
| 2.6 Modelling of real freeform surfaces | 23 |
| 2.6.1 Power spectral density | 24 |
| 2.6.2 Description of low spatial frequency errors | 25 |
| 2.6.3 Description of deterministic ripples | 27 |
| 2.6.4 Description of random errors | 27 |
| 2.6.5 System performance criteria | 29 |
| 2.7 Open questions | 31 |
| 3 General Solution for Obscuration Constraints | 33 |
| 3.1 Construction of error functions | 33 |
| 3.2 Workflow for optimization | 38 |
| 3.3 Obscuration elimination of three-mirror Yolo telescope | 40 |
| 3.4 Two-mirror Yolo telescope design without astigmatism | 42 |
| 4 Quality Assessment of Chromatic Aberrations in Refractive Freeform Systems | 48 |
| 4.1 Reference plane and reference axis | 48 |
| 4.2 Definition of the generalized chromatic aberrations | 50 |
| 4.3 Ray-based calculation method | 52 |

Contents

| | | |
|--------------------------|--|-----|
| 4.3.1 | Chromatic image variation | 52 |
| 4.3.2 | Transverse chromatic aberration | 55 |
| 4.4 | Wavefront-based calculation method | 57 |
| 4.4.1 | Real wavefront and reference sphere | 57 |
| 4.4.2 | Additivity of phase | 59 |
| 4.4.3 | Fitting of the chromatic aberrations | 59 |
| 4.5 | Influence of chromatic pupil aberration | 60 |
| 4.6 | Examples | 62 |
| 4.6.1 | Double Gauss system | 62 |
| 4.6.2 | Stereomicroscope | 64 |
| 4.6.3 | Anamorphic photographic lens | 68 |
| 4.6.4 | HMD system..... | 73 |
| 5 | Tolerancing of Freeform System | 78 |
| 5.1 | Autocorrelation function of unwrapped phase pattern and its characterization ... | 78 |
| 5.2 | Parametrization of real freeform surfaces | 84 |
| 5.3 | Modelling a synthetic surface..... | 87 |
| 5.4 | Example of tolerancing a freeform system..... | 89 |
| 6 | Conclusions..... | 98 |
| Appendix | A..... | 100 |
| Appendix | B..... | 101 |
| References | | 102 |
| List of Figures | | 107 |
| List of Tables | | 113 |
| List of Abbreviations | | 114 |
| List of Symbols | | 116 |
| Acknowledgement | | 120 |
| Ehrenwörtliche Erklärung | | 121 |
| Publications | | 122 |
| Curriculum Vitae..... | | 123 |

1 Motivation

Nowadays, modern optical systems play an important role in our life. Scientists use high precision telescopes to detect the deep universe; people use cameras to record the beautiful moments; augmented reality (AR) and virtual reality (VR) devices break the gap between the virtual and real worlds. Looking back through the history, the first optical system, which was a two-element microscope, was invented by the father and son team of Hans and Zacharias Janssen in 1590 [1-1]. Then, with the collaboration among Zeiss, Schott and Abbe, the optical industry became successful in the late of the 19th century [1-1]. Starting from the characteristic function of Hamilton [1-2] and eikonal function [1-3], the aberration theory has been developed, which helps optical designers understand the optical system and turns imaging optics into a science.

In the early years before 1950s, optical systems were designed with surfaces aligned on the same axis. The light propagation through optical systems is described by the geometrical model, in which finite rays are traced through the whole system and forms the geometrical wavefront in the image space. The difference between the real wavefront and ideal wavefront is the aberration in an optical system. The aberrations are expanded mathematically into polynomials of different orders [1-4]. The aberration coefficients of the primary aberrations in the rotationally symmetric system were derived analytically by Seidel [1-5]. The so-called Seidel aberrations show the surface contributions to each type of aberration, and the formulae contain the physical quantities of the paraxial marginal ray (MR) and paraxial chief ray (CR) only. Therefore, it reduced the workload of the optical designers especially before the invention of computer, and becomes a powerful tool in analyzing what happens in optical systems. The development of materials in this period cannot be neglected as well. The newly glasses with different refractive indices and Abbe numbers were invented [1-6], which gave optical designers more choices. As a result, optical systems were developed with more elements to get better image quality. Correction strategies are more oriented under the instruction of aberration theory. However, the structure of the system is still limited to rotational symmetry. The drawback is seen in telescopes, in which the central obscuration is unavoidable [1-7, 1-8].

From the mid of the 20th century, technology develops rapidly that provides more possibilities in design of optical systems. With the improvement of the manufacturability of the optical elements, more types of optical surfaces are applied to optical design. For

example, cylindrical and toroidal surface are helpful in designing anamorphic systems [1-9, 1-10], aspherical and freeform surfaces are broadly used to enhance the system performance for the reason that they are superior in controlling the ray path locally [1-11, 1-12]. In addition, tilting and decentering of the optical elements provide more degrees of freedom in optimization to achieve special requirements. For example, mirrors are tilted in off-axis reflective systems to overcome obscuration, Alvarez plates are decentered to get dynamic focal lengths [1-13, 1-14], the objective lens in the stereomicroscope is decentered to reach double-channel stereo imaging. Both the actions of adopting asymmetric surfaces and moving optical elements apart from the common axis break the symmetry of optical systems, and thus bring challenges to the traditional aberration theory, optical design and tolerancing.

When optimizing the off-axis telescope systems, it is not trivial to find out the proper tilting angles of the mirrors. On the one hand, the tilting angles are related to the image quality; one the other hand, the mirrors should not block the ray path. Xu *et al.* has introduced an error function to describe all the obscuration cases in plane-symmetric optical systems and embedded the error function into optimization algorithm to search for the obscuration-free structure automatically [1-15]. However, his method is not applicable to non-rotationally symmetric optical systems, in which no symmetric plane exists.

Nodal aberration theory was developed by Thompson in 1980 to describe the aberrations in off-axis optical systems [1-16]. Unlike the rotationally symmetric case, the fields and pupil coordinates are represented in the vectorial form. The theory expands the wavefront analytically to the 6th order and gives insights into the field-dependent aberrations by introducing the Nodal points. But his theory is working monochromatically. Cao *et al.* attempted to consider the chromatic aberrations in plane-symmetric optical systems [1-17]. Nevertheless, the total chromatic aberrations are calculated only for the fields in the symmetry plane. The surface contributions cannot be resolved neither.

Tolerancing of freeform surface is a practical topic. The surface errors can be classified into low spatial frequency (LSF) errors, mid spatial frequency (MSF) errors, and high spatial frequency (HSF) errors. Although there are many papers discussing about how to model the errors in a specific frequency range [1-18]-[1-22], none of them connects all the errors together to establish a real freeform surface, which is ready for tolerancing.

The purpose of this thesis is to propose a generalized method to eliminate obscuration automatically and a generalized theory to describe, calculate and analyze chromatic aberrations in non-symmetrical optical systems. With this work, the method and the theory

are consistent and can be applied in arbitrary optical systems regardless of surface type and system structure. In addition, a comprehensive model to simulate freeform surface errors in all the spatial frequency ranges is presented. Based on this model, many synthetic surfaces can be generated to predict the real freeform surfaces.

Chapter 2 gives a short overview of the state-of-the-art freeform optical systems and aberration theory. It starts with the development of the optical surfaces, especially the representations of the aspherical and freeform surfaces are introduced. Then, the optical systems are classified according to their geometry. Several reflective and refractive optical systems are given as examples. The evolution of the surface imaging equations and aberration theories are introduced with the lack of symmetry step-by-step. The obscuration issue in the reflective optical system is discussed and the current existing obscuration elimination method is reviewed. In the last section, the models to describe the freeform surface errors in different frequency ranges are discussed, which are the foundation of the comprehensive model.

In chapter 3, the automatic obscuration elimination algorithm used in non-rotationally symmetric reflective optical systems is developed. An error function, which contains four parts to fully represent all kinds of obscurations, is described in detail. Then the workflow to remove obscuration by global optimization is proposed. Two examples are given to show the feasibility of the algorithm. Especially in the second example, the obscuration is eliminated together with astigmatism to obtain an initial system for further optimization.

Chapter 4 focuses on the generalized chromatic aberrations in non-rotationally symmetric refractive optical systems. The generalized chromatic aberrations are defined and compared with the traditional definitions in the rotationally symmetric systems after figuring out the reference plane and reference axis. The surface-by-surface summation of the generalized chromatic aberrations for a single field are calculated by both ray-based and wavefront-based methods in an analytical-numerical hybrid manner. The aberration distribution for the whole field of view (FOV) can be seen by the full-field-display. The impact of chromatic pupil aberration on the calculation accuracy is discussed, and the calculation methods are modified accordingly. The first example of double Gauss system demonstrates the consistency of the new methods with the Seidel aberrations in the first order approximation. The second example shows the consistency between the ray-based and wavefront-based methods. The influence of freeform on the generalized chromatic aberrations is discussed in the third example. The fourth example evaluates the behavior of the intrinsic and induced chromatic aberrations in a non-rotationally symmetric optical system.

Chapter 5 is about the tolerancing of freeform surface. Both the unwrapped phase pattern and the power spectral density (PSD) of the surface errors in the frequency domain are computed. The difference between the deterministic and statistic errors is compared by characterizing the autocorrelation function of the unwrapped phase pattern to find out the boundary frequency between the two kinds of errors. After determining the boundary frequency, the LSF errors are generated by tolerancing the Zernike coefficients with the given root-mean-square (RMS) surface error. The PSD of the LSF errors is plotted in the log-log scale. The linear curve is fitted and extended to the HSF region. The statistic HSF errors are generated according to the area under the PSD in the HSF region. The MSF errors are added to the surface map as the cosine waves. In the above approach, a synthetic freeform surface is simulated and imported back into the optical design software for analysis.

Finally, chapter 6 addresses the conclusion of this thesis and outlook of further work.

2 Theoretical Introduction

As shown in Fig. 2-1, the geometry of optical surfaces and systems can be classified into rotational symmetry, double-plane symmetry, plane symmetry, and non-symmetry. The first half of this chapter reviews the development of optical surfaces and systems, surface imaging equations, and aberrations theories regarding the reduction of symmetry over the last decades. Then, the practical issues are introduced, for example, the obscuration in mirror systems and its solution of the auto-optimization algorithm, the modelling of a real freeform surface with a general profile which is manufactured by the diamond turning machine because of its robustness, low cost and efficiency. The last section discusses the some important questions in the development of freeform optical systems from design to production.

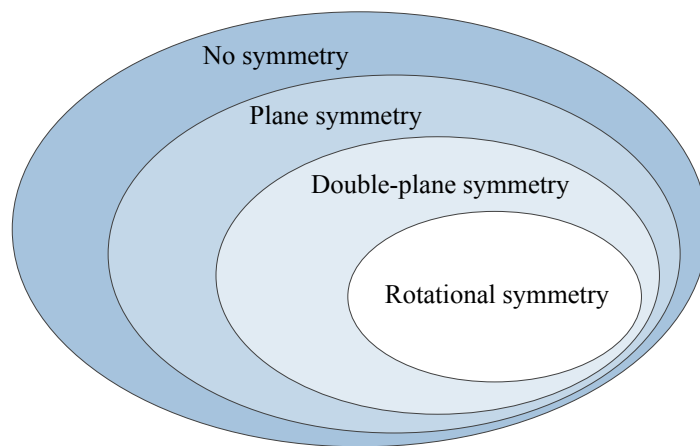


Fig. 2-1 Geometry of optical surfaces and systems. The surface shape and the bending of the optical axis together determine the geometry of an optical system.

2.1 Optical surfaces

2.1.1 Basic shapes

The basic shape of an optical surface is described by the second order mathematical function. The formula in Cartesian coordinates is written as

$$z_{\text{basic}}(x, y) = \frac{c_x x^2 + c_y y^2}{1 + \sqrt{1 - (1 + \kappa_x) c_x^2 x^2 - (1 + \kappa_y) c_y^2 y^2}} \quad (2.1)$$

where x and y are the coordinates on the surface, z is the surface sag, the radii of curvature c_x and c_y and the conic constants κ_x and κ_y are specified in the two perpendicular cross sections, respectively. The relations between the value of conic constant and the surface shape in the corresponding cross section are summarized in Table 2-1. If $c_x = c_y$ and $\kappa_x = \kappa_y$

are fulfilled at the same time, the surface is rotationally symmetric and can be written in the polar coordinates as

$$z_{\text{basic}}(r) = \frac{cr^2}{1 + \sqrt{1 - (1 + \kappa)c^2r^2}}. \quad (2.2)$$

Especially, the surface is called spherical surface when $\kappa=0$. If the radius of curvature and the conic constant in one cross section are both zero, the surface is called cylindrical surface. A biconic surface is obtained if the non-zero radii of curvature and conic constants in both cross sections are different. Both cylindrical and biconic surfaces are double-plane symmetric.

| Conic constant | Surface shape |
|-------------------|-----------------|
| $\kappa < -1$ | Hyperbola |
| $\kappa = -1$ | Parabola |
| $-1 < \kappa < 0$ | Prolate ellipse |
| $\kappa = 0$ | Sphere |
| $\kappa > 0$ | Oblate ellipse |

Table 2-1 Conic constant and surface shape.

2.1.2 Aspheres

The second order basic shape functions describe the dominant sag change on the optical surface, and it is possible to add higher order polynomial terms as the deviation to the basic shape to describe the local change on the surface profile, which is helpful to correct aberrations with its local curvature to bend the rays locally.

The simplest surface with polynomial deformation is the aspherical surface, which keeps the geometry of rotational symmetry. Traditionally, the aspherical deviation can be represented by a Taylor expansion of even powers as

$$z(r) = \frac{cr^2}{1 + \sqrt{1 - (1 + \kappa)c^2r^2}} + \sum_{m=1}^M a_{2m+2}r^{2m+2} \quad (2.3)$$

where a_{2m+2} are the expansion coefficients, m starting from 1 shows that the Taylor expansion starts from the 4th order, M is the maximum term number of the expansions. In practice, up to seven expansion terms are used. More than 10 terms will cause cancelling effect among the orders and strong oscillation of the surface profile at the edge [2-1, 2-2].

In order to overcome the above drawback of Taylor expansion, two kinds of orthogonal polynomial sets were introduced by Forbes [2-2] to describe aspherical surfaces. According

to different orthogonal conditions, the Forbes aspheres can be classified into strong aspheres and mild aspheres.

The analytical equation of the strong asphere is written as

$$z(r) = \frac{cr^2}{1 + \sqrt{1 - (1 + \kappa)c^2r^2}} + \bar{r}^4 \sum_{m=0}^M a_m Q_m^{\text{con}}(\bar{r}^2) \quad (2.4)$$

where $\bar{r} = \frac{r}{r_{\text{norm}}}$ is normalized by the radius r_{norm} . Let $x = \bar{r}^2$, then $x^2 Q_m^{\text{con}}(x)$ is a polynomial set in which the polynomial terms are orthogonal along the z -axis.

The analytical equation of the mild asphere is written as

$$z(r) = \frac{c_{\text{bfs}}r^2}{1 + \sqrt{1 - c_{\text{bfs}}^2r^2}} + \frac{\bar{r}^2(1 - \bar{r}^2)}{\sqrt{1 - c_{\text{bfs}}^2r^2}} \sum_{m=0}^M a_m Q_m^{\text{bfs}}(\bar{r}^2) \quad (2.5)$$

in which the basic shape is the best fitted sphere with the radius of curvature as c_{bfs} . The polynomial set $x^2(1 - x^2)Q_m^{\text{bfs}}(x^2)$ is slope orthogonal. The polynomial terms are divided by a projection factor of

$$\cos \phi_{\text{pr}} = \sqrt{1 - c_{\text{bfs}}^2r^2} \quad (2.6)$$

to project the polynomials to the direction along the z -axis.

For the polynomial set $x^2 Q_m^{\text{con}}(x)$, all the terms equal 0 at the center and 1 on the boundary. Therefore, the deviation of an asphere to the conic basic shape cannot be controlled at the edge. That is the reason why it is called strong asphere. For the polynomial set $x^2(1 - x^2)Q_m^{\text{bfs}}(x^2)$, all the terms equal 0 at both center and edge to suppress the aspherical deviation. Therefore, such an asphere described by $Q_m^{\text{bfs}}(x^2)$ is called mild asphere. The differences between strong and mild aspheres are graphically illustrated in Fig. 2-2.

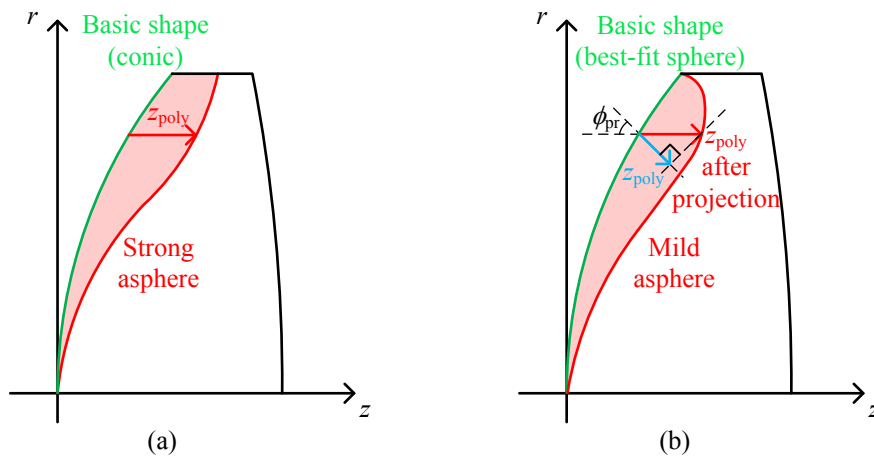


Fig. 2-2 Comparison between (a) strong and (b) mild aspheres regarding basic shape and orthogonality.

2.1.3 Freeforms

If the restriction of rotational symmetry is removed, general freeform surfaces can be obtained by the combination of varying basic shapes and polynomial sets. The basic shape can be either rotationally symmetric, which is described in polar coordinates, or double-plane symmetric, which is described in Cartesian coordinates. The commonly used polynomial sets are monomials, Chebyshev polynomials [2-3], Zernike polynomials, Q-polynomials [2-4], and newly invented A-polynomials [2-5].

The freeform surfaces investigated in this thesis are described by the Zernike Fringe polynomials, which are directly related to aberration types and can be sorted in 1D or 2D indices, as

$$z(r, \varphi) = z_{\text{basic}} + \begin{cases} \sum_{n=0}^N \sum_{m=-n}^n a_{n,m} Z_m^n(\bar{r}, \varphi), & \text{2D numbering} \\ \sum_{k=1}^K a_k Z_k(\bar{r}, \varphi), & \text{1D numbering} \end{cases} \quad (2.7)$$

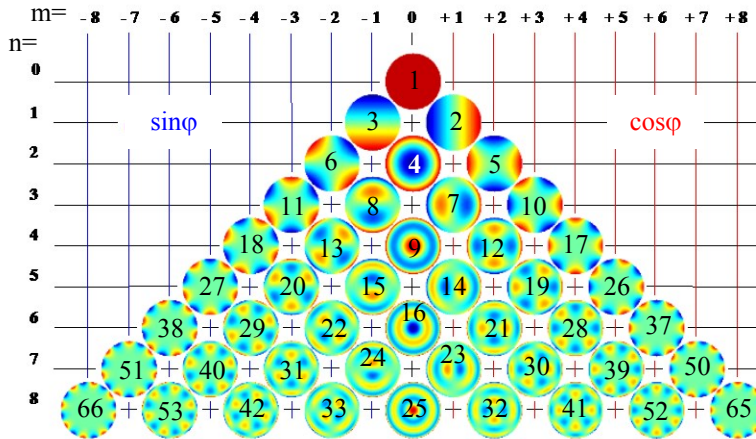


Fig. 2-3 Zernike Fringe polynomials in 2D sorting (m and n numbering on the edge) and 1D sorting (numbering on the pattern).

The patterns of Zernike Fringe polynomials normalized on the unit circle are plotted in Fig. 2-3 with two numbering ways. By selecting proper polynomial terms with double-plane symmetry or plane symmetry to describe a freeform surface, one can get the freeform surface with the corresponding symmetry.

2.2 Classification of optical systems

In the long history, even nowadays, rotationally symmetric optical systems are dominant because they are sufficient in most applications and relatively easy to produce. Nevertheless,

non-rotationally symmetric optical systems are more and more popular to achieve better system performance and special requirements of compactness or application adapted geometries. In this section, modern optical systems are classified into different groups according to their geometry.

The family of rotationally symmetric optical systems is broad. Some of them are listed in Table 2-2. Traditional refractive optical systems, such as photographic lenses, microscope objective lenses, endoscopes, etc., belong to this family. The complexity of the system depends on the requirements of the system's performance. Cell phone camera systems are also rotationally symmetric. However, restricted by size and weight, the number of surfaces in these systems is limited, and plastic components with strong dispersion are used. Therefore, aspherical surfaces are introduced for better correction. The reflecting telescope systems designed in the early years of the 20's century were rotationally symmetric as well. The conic mirrors are confocal to form a perfect image for one field point. Since the field angles are normally small in a telescope, this kind of design can serve as a good starting point for further correction. However, the incoming beam is partly blocked by the mirrors (only the yellow shaded part can go through the example system in Table 2-2), the systems are unavoidably suffering from central obscuration. Another type of rotationally symmetric system is called catadioptric system, which contains both lenses and mirrors. Usually, a simple achromat with small focal power is inserted in front of the reflecting telescope to co-

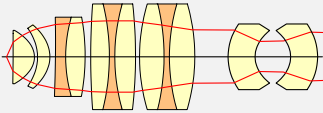
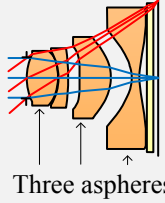
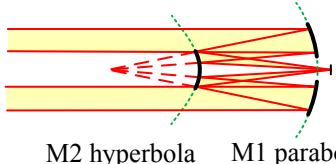
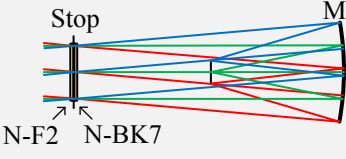
| Types | Systems |
|--|---|
| <p style="text-align: center;">Refractive</p> | <p>Microscope objective lens </p> <p>Cell phone camera lens  Three aspheres</p> |
| | <p>Cassegrain telescope  M2 hyperbola M1 parabola</p> |
| <p style="text-align: center;">Catadioptric</p> | <p>Catadioptric telescope  Stop M1 N-F2 N-BK7</p> |

Table 2-2 Some rotationally symmetric optical systems [2-6, 2-7].

rect spherical aberration. The refractive surfaces can also be made aspherical.

The optical elements in double-plane symmetric optical systems are still aligned; only at least one spherical surface is substituted by a cylindrical surface, biconic surface, or freeform surface represented by double-plane symmetric polynomial terms. In principle, double-plane symmetric systems have different focal power in the two perpendicular cross sections. Three double-plane symmetric systems are shown in Fig. 2-4. The laser shaping system shown in Fig. 2-4 (a) transforms the anastigmatic output laser beam into a collimated astigmatic ray bundle by combining a spherical lens and a cylindrical Galilean telescope. The example in Fig. 2-4 (b) is an anamorphic photographic lens, which captures an image or projects the cinemascope film of a large aspect ratio. Two cemented cylindrical lens groups are put in front of spherical lenses to realize different magnifications in the two cross sections. In scan systems, the scan mirror locates at the front focal point to achieve a telecentric chief ray in the image space, if this is required. However, it works for one dimension only. With the help of the two cylindrical lenses in Fig. 2-4 (c), the focal points in the two cross sections are separated. Two mirrors are located at the split focal points and rotate independently in two directions to fulfill the telecentric conditions in two dimensions.

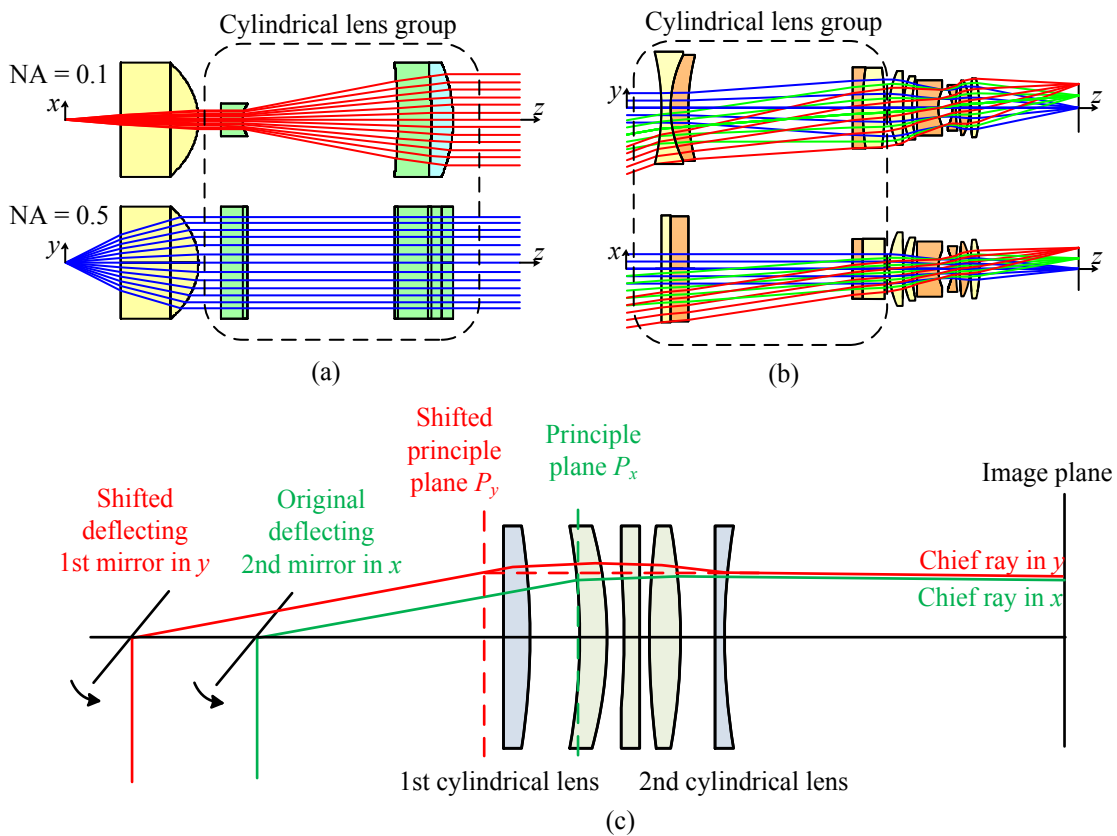


Fig. 2-4 Three examples of double-plane symmetric systems: (a) laser shaping system [2-8], (b) anamorphic photographic lens [2-9], (c) 2D telecentric scan system [1-11].

Plane-symmetric optical systems have only one symmetric plane, which is conventionally defined as the tangential plane. The optical elements can either be aligned on the same axis with the application of a plane-symmetric freeform surface to break the symmetry, or tilt or decenter around a common axis. The optical axis ray (OAR), which is the central ray of the central field, is folded in the tangential plane. The layouts of some typical plane-symmetric

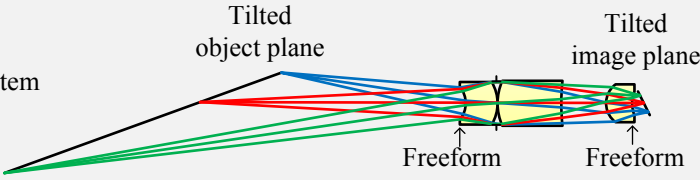
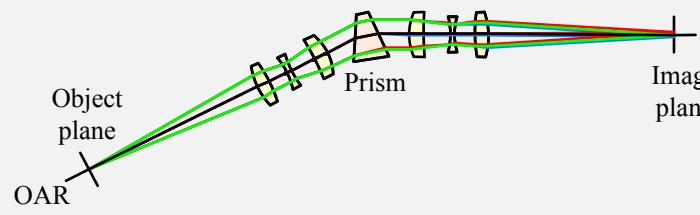
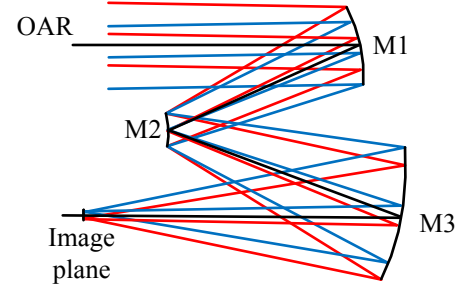
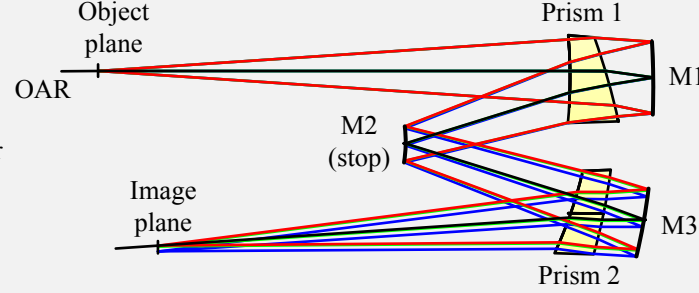
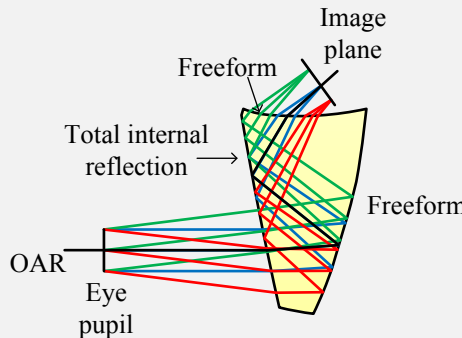
| Types | Systems |
|-----------------------------|--|
| <p>Refractive</p> | <p>Scheimpflug system</p>  <p>Tilted object plane</p> <p>Tilted image plane</p> <p>Freeform</p> <p>Freeform</p> <p>Traditional spectrometer</p>  <p>Object plane</p> <p>Image plane</p> <p>OAR</p> <p>Prism</p> |
| | <p>Three mirror anastigmat (TMA) system</p>  <p>OAR</p> <p>M1</p> <p>M2</p> <p>M3</p> <p>Image plane</p> |
| <p>Cata-dioptric</p> | <p>Offner spectrometer</p>  <p>Object plane</p> <p>Image plane</p> <p>OAR</p> <p>Prism 1</p> <p>M1</p> <p>M2 (stop)</p> <p>M3</p> <p>Prism 2</p> <p>HMD</p>  <p>Image plane</p> <p>Freeform</p> <p>Freeform</p> <p>OAR</p> <p>Eye pupil</p> <p>Total internal reflection</p> |

Table 2-3 Layouts of some plane symmetric optical systems in the tangential plane [2-6, 2-10, 2-11]. The bended OAR is drawn in each layout.

systems are shown in Table 2-3. Scheimpflug system is a special refractive plane-symmetric system, which images a tilted object plane sharply to a tilted image plane. The lenses can be purely spherical, but the whole geometry of the system usually is plane-symmetric. In general, the surfaces can be tilted or substituted by plane-symmetric freeform surfaces for better correction. The structure of a spectrometer is also plane-symmetric, which separates the spectral components of the incoming light in the lateral direction. The system can be purely refractive or folded as an Offner setup with the help of mirrors. The benefit of the Offner setup is that the ray path is doubled so that the separation between the wavelengths is more significant. Another refractive-reflective plane symmetric system is called head-mounted-display (HMD). The folded structure reduces the size of the wearable device, and the freeform surface ensures high image quality in the case of small F-numbers and large FOV. A plane symmetric mirror system, in which the mirrors are tilted around one common axis, is preferred in the design of the telescope system because the central obscuration in the reflective telescope can be fully removed to improve the contrast.

A Yolo telescope is a kind of optical system without any symmetry, as shown in Fig. 2-5. The mirrors rotate around both x - and y -axes. The original design has two mirrors. An extended Yolo telescope with three mirrors is also existing [2-12].

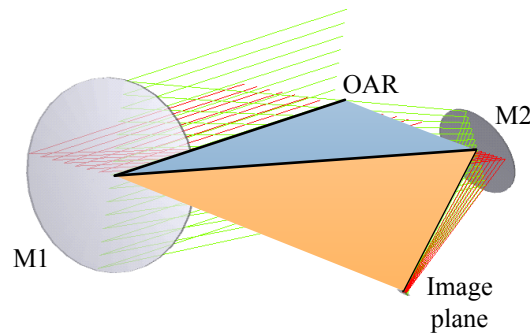


Fig. 2-5 Yolo telescope with two mirrors. The OAR is folded in three-dimensional (3D) space with three segments. Two pairs of the neighboring segments construct two planes in blue and orange to indicate the 3D folding more clearly.

2.3 Single surface imaging equations

In the paraxial region of a spherical surface, the imaging equation of a single field point is governed by

$$\frac{n'}{l'} - \frac{n}{l} = \frac{n' - n}{R} = \frac{1}{f} \quad (2.8)$$

where l and l' are the object and image distances, n and n' are the refractive indices in the object and image spaces, R is the radius of curvature of the spherical surface, f is the focal length.

For an off-axis field away from the paraxial region, astigmatism cannot be neglected. The effective radius of curvature of a spherical surface depends on the azimuthal angle of the ray fan, which means that the tangential and sagittal rays focus on different points on the chief ray as shown in Fig. 2-6. At the tangential focal point, the sagittal rays form the tangential line in the sagittal plane, and vice versa. The image distances along the chief ray in the tangential and sagittal cross sections can be reached by the Coddington equations as

$$\frac{n' \cos^2 I'}{t'} - \frac{n \cos^2 I}{t} = \frac{n' \cos I' - n \cos I}{R} \quad (2.9)$$

$$\frac{n'}{s'} - \frac{n}{s} = \frac{n' \cos I' - n \cos I}{R} \quad (2.10)$$

where t, t' and s, s' are the distances along the chief ray in the corresponding cross sections in the object and image spaces, I and I' are the incidence and refraction angle of the chief ray. When the incidence angle $I=0$, the equations are degraded to Eq. (2.8).

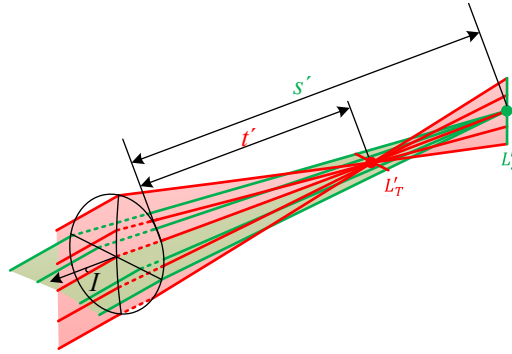


Fig. 2-6 Separation of tangential and sagittal focal points due to the inclined incidence angle I . L'_T and L'_S denote the tangential and sagittal focal lines.

The imaging equation of a freeform surface is more complicated because the radii of curvature are locally variant. Considering an astigmatic ray bundle going through a freeform surface with an infinite small numerical aperture (NA) as shown in Fig. 2-7, the local incidence plane (yellow) on the surface is determined by the incident and refracted chief rays. The local surface coordinate system (x_s, y_s, z_s) is constructed with the principle that the y_s - z_s plane coincides with the incidence plane and the z_s -axis is along the surface normal at the chief ray intersection point on the surface. The surface illuminated by the light can be described locally around the chief ray by the XY polynomials. Since the light cone aperture is considerably small, it is sufficient to approximate the surface segment until the second order expansion as

$$z_s = \frac{1}{2}(c_{02}x_s^2 + c_{20}y_s^2 + c_{11}x_s y_s) = \frac{1}{2}[(c_{\max} \cos^2 \eta + c_{\min} \sin^2 \eta)x_s^2 + (c_{\max} \sin^2 \eta + c_{\min} \cos^2 \eta)y_s^2 + (c_{\max} - c_{\min})x_s y_s \sin 2\eta] \quad (2.11)$$

2 Theoretical Introduction

where c_{02} , c_{20} , c_{11} are the polynomial coefficients. In this way, the local surface segment is approximated by a toroidal surface with the maximum and minimum principal curvatures c_{\max} and c_{\min} (not drawn), which are perpendicular to each other. The azimuthal angle between the x_s -axis and the direction of the maximum principal curvature is η .

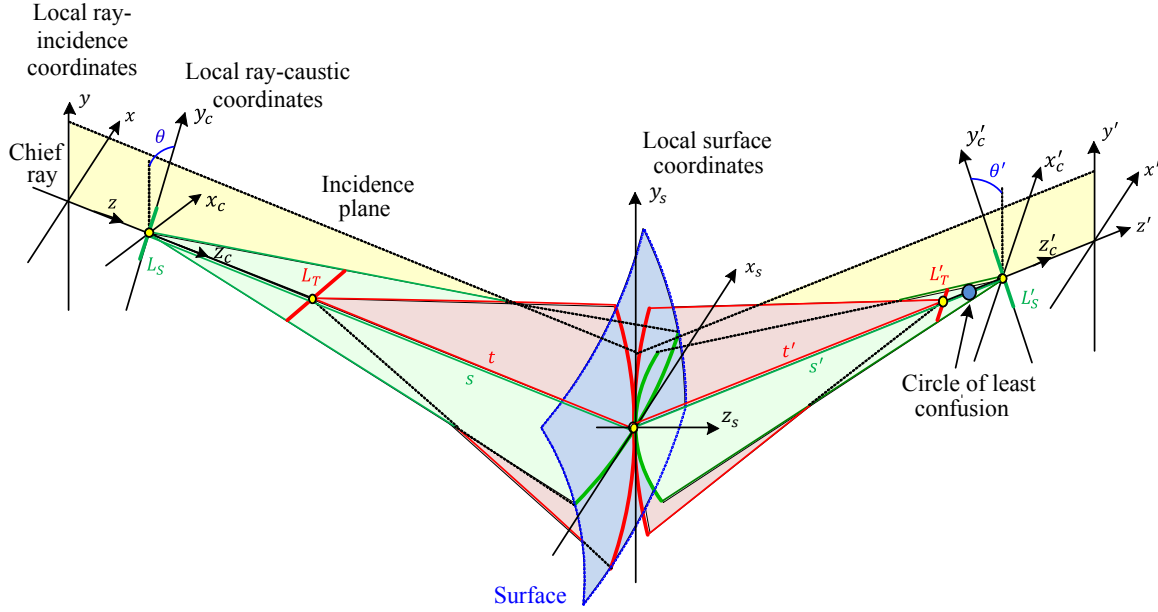


Fig. 2-7 An astigmatic ray bundle going through a freeform surface with an infinite small numerical aperture. The surface segment illuminated by the ray bundle is approximated by a toroidal surface.

Two types of local coordinate systems, which consider the azimuth of the ray bundle, have to be introduced. The first is the local ray-incidence coordinate system (x, y, z) in the object space, and accordingly, (x', y', z') in the image space. They all have the y - z plane coinciding with the incidence plane and the z -axis along the chief ray. The second type is called the local ray-caustic coordinate system, which indicates the azimuth of the focal lines. It is denoted by (x_c, y_c, z_c) in the object space. The x_c - and y_c -axes are in the direction of the tangential and sagittal focal lines, respectively. The z_c -axis is also along the chief ray. The local ray-caustic coordinates (x'_c, y'_c, z'_c) in the image space are constructed in the same way with the focal lines in the image space. Now, both the surface segment and the ray bundle are anisotropic, and it may happen that the tangential cross section of the ray bundle does not coincide with the incidence plane. The azimuthal angles between the ray-incidence and ray-caustic coordinate systems in the object and image space are θ and θ' . The image distances t' and s' along the chief ray and θ' can be calculated by the generalized Coddington equations [2-13]

$$(n \cos I - n' \cos I')(c_{\max} \cos^2 \eta + c_{\min} \sin^2 \eta) = n \left(\frac{\cos^2 \theta}{s} + \frac{\sin^2 \theta}{t} \right) - n' \left(\frac{\cos^2 \theta}{s} + \frac{\sin^2 \theta}{t} \right) \quad (2.12)$$

$$(n \cos I - n' \cos I')(c_{\max} - c_{\min}) \sin 2\eta = n \cos I \left(\frac{1}{s} - \frac{1}{t} \right) \sin 2\theta - n' \cos I' \left(\frac{1}{s'} - \frac{1}{t'} \right) \sin 2\theta' \quad (2.13)$$

$$(n \cos I - n' \cos I')(c_{\max} \sin^2 \eta + c_{\min} \cos^2 \eta) = n \cos^2 I \left(\frac{\sin^2 \theta}{s} + \frac{\cos^2 \theta}{t} \right) - n' \cos^2 I' \left(\frac{\sin^2 \theta'}{s'} + \frac{\cos^2 \theta'}{t'} \right) \quad (2.14)$$

2.4 Aberration theories

Aberrations are unavoidable in real optical systems as a result of nonlinearities for large ray-incidence angles, except for some systems under special imaging conditions, for example, the confocal conic mirror telescope introduced in section 2.2. In order to evaluate and improve the system performance more systematically with guidance, it is important to describe the aberrations quantitatively.

In general, aberrations can be represented as wave aberrations and geometrical aberrations within the scope of geometrical optics. Moreover, the geometrical aberrations can be further specified into transverse aberrations and longitudinal aberrations. The real and ideal ray paths from the exit pupil to the Gaussian image plane can be seen in Fig. 2-8. The ideal and real ray fans construct the reference sphere with radius R_{ref} and the real wavefront, respectively, in the exit pupil. The wave aberration W is the difference between the reference sphere and real wavefront, which is calculated by the optical path difference (OPD) between the ideal and real rays as

$$W = \frac{n' \cdot \text{OPD}}{\lambda} \quad (2.15)$$

where n' is the refractive index in the image space, λ is the wavelength under consideration. The transverse aberrations $\Delta x'$ and $\Delta y'$ in x - and y -directions represent the difference between the ideal and real ray intersection points in the image plane. The longitudinal aberration $\Delta s'$ measures the difference between the ideal and real ray intersection points with the optical axis.

The relations among the three aberration representations in the small angle approximation are

$$\Delta x' = -\frac{R_{\text{ref}}}{n'} \frac{\partial W(x_p, y_p)}{\partial x_p} \quad (2.16)$$

$$\Delta y' = -\frac{R_{\text{ref}}}{n'} \frac{\partial W(x_p, y_p)}{\partial y_p} \quad (2.17)$$

$$\Delta s' = \frac{R_{\text{ref}}}{y_p} \Delta y' \quad (2.18)$$

where x_p, y_p are coordinates on the exit pupil. Considering the sign convention, the wave aberrations have the opposite sign to both transverse and longitudinal aberrations.

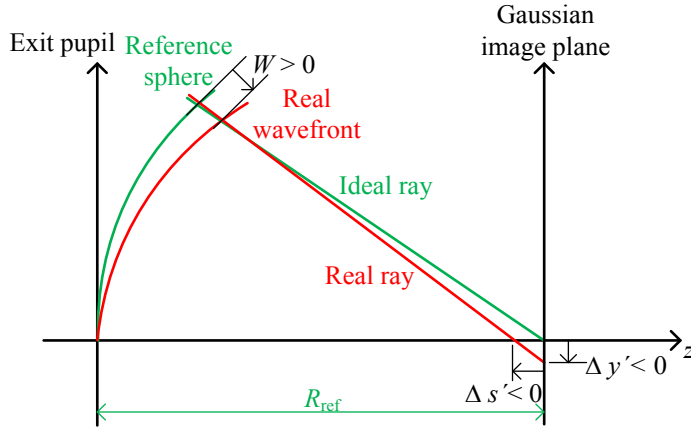


Fig. 2-8 Relations between wave aberration, transverse aberration, and longitudinal aberration.

2.4.1 Aberration theories in rotationally symmetric optical systems

Both wave aberrations and geometrical aberrations can be described by polynomial expansions. It can be seen from Eq. (2.16) to (2.18) that the wave aberrations are one order higher than the transverse aberrations, and the transverse aberrations are one order higher than the longitudinal aberrations. In general, wave aberrations can be expanded by Taylor expansion as

$$W(H, \rho, \varphi) = \sum_{k,l,m} W_{klm} H^k \rho^l \cos^m \varphi \quad (2.19)$$

where H is the field height in the image plane, ρ and φ are the radial height and azimuthal angle of the ray intersection point in the exit pupil. For a single field, the 2D wavefront deviation in the exit pupil is calculated by tracing the rays sampled in the pupil or measured by interferometry. It is more meaningful to select the exit pupil, but it is also possible to use the entrance pupil as Zemax does. The 2D wavefront map can be fitted by the Zernike polynomials as

$$W(\rho, \varphi) = \sum_k a_k Z_k(\rho, \varphi) \quad (2.20)$$

which includes all orders of monochromatic aberrations.

The transverse aberration expressed by Taylor expansion is written as

$$\Delta y'(H, \rho, \varphi) = \sum_{k,l,m} W_{klm} H^k \rho^l \cos^m \varphi. \quad (2.21)$$

There are five monochromatic primary aberrations: spherical aberration, coma, astigmatism, field curvature, and distortion. The primary aberrations represented as wave aberrations and transverse aberrations are of 4th and 3rd order respectively. The two primary chromatic aberrations are called longitudinal chromatic aberration, which is the dispersion of marginal ray along the optical axis, and transverse chromatic aberration, which is the dispersion of chief ray along the transverse direction in the Gaussian image plane, as illustrated in Fig. 2-9. Seidel derived the expression of the seven primary aberrations in the rotationally symmetric systems analytically. The formulae contain only the physical quantities of the paraxial marginal ray and paraxial chief ray and the derivative of the refractive index in case of the two chromatic aberrations. In addition, the Seidel aberrations can be written as the sum of surface contributions.

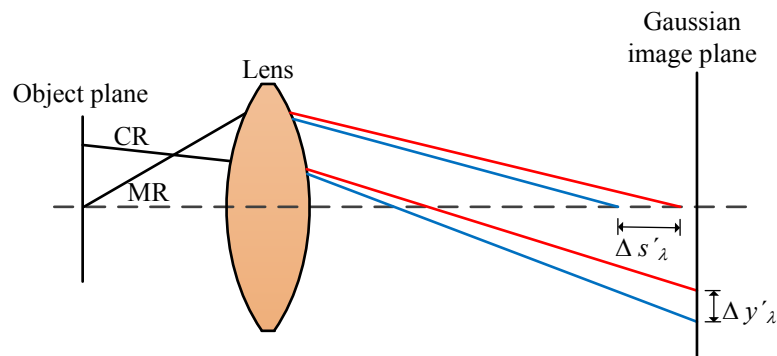


Fig. 2-9 Longitudinal chromatic aberration $\Delta s'_\lambda$ and transverse chromatic aberration $\Delta y'_\lambda$ between two wavelengths (blue and red).

The properties of the Seidel aberrations and the wave aberrations represented by Zernike polynomials are compared in Table 2-4 to summarize the above discussion.

| | Seidel aberrations | Zernike polynomials |
|-------------|--|--|
| Pros | Surface contributions, Monochromatic and chromatic aberrations, The field dependence is included, Only two paraxial rays are required | Can be measured by interferometry directly, Higher order aberrations can be seen |
| Cons | Only primary aberrations, Only rotationally symmetric systems | May have sampling problems, Only one field point, Restricted to circular pupils, Only monochromatic aberrations |

Table 2-4 Comparison between Seidel aberrations and wave aberrations represented by Zernike polynomials.

2.4.2 Aberration theories in non-rotationally symmetric optical systems

To tackle the aberrations in non-rotationally symmetric optical systems, Thompson *et al.* [1-16, 2-14, 2-15] have extended the wave aberration function into the vectorial form as

$$W(\vec{H}, \vec{\rho}) = \sum_{k,l,m} W_{klm} (\vec{H} \cdot \vec{H})^k (\vec{\rho} \cdot \vec{\rho})^l (\vec{H} \cdot \vec{\rho})^m \quad (2.22)$$

in which the scalar field and pupil height in Eq. (2.19) are substituted by vectors. The reference axis is defined as the folded OAR, and the intersection of OAR in the image plane is chosen to be the image center. Moreover, analogous to the concept of the paraxial region, the neighborhood region around the OAR is called the parbasal region and serves as a reference, although the OAR suffers from ray aberrations.

As demonstrated [1-16], both tilt and decenter of the surface can be viewed as tilt and introduce a field shift vector $\vec{\sigma}_i$, which results in the displacement of the field center in the image plane as shown in Fig. 2-10. The effective field vector \vec{H}_{Ai} for the surface i after considering the tilt effect is then written as

$$\vec{H}_{Ai} = \vec{H} - \vec{\sigma}_i. \quad (2.23)$$

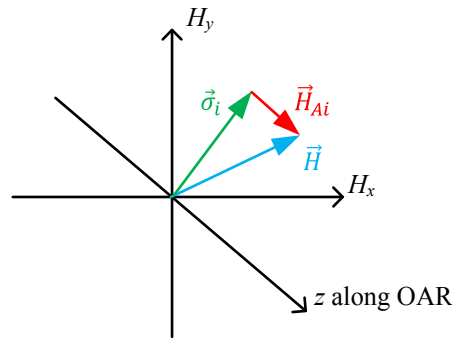


Fig. 2-10 Displacement of the effective field center of the surface i in the image plane [1-16].

The advantage of the vectorial wave aberration expansion with the tilt term is that the nodal points at which the primary aberrations have zero values can be derived. Therefore, it is also called Nodal aberration theory (NAT). This approach initially requires circular symmetric surfaces, later a more complicated extension to freeforms was proposed [2-16]. It is also a great advantage to plot the aberration coefficients for each type of aberrations as the full-field display to analyze the system behavior across the whole image plane. Nevertheless, the chromatic aberrations are of no concern in the NAT.

Cao *et al.* presented analytical solutions to the chromatic aberrations in the tangential plane of plane-symmetric optical systems [1-17]. The analytical formulae were derived by differentiating the Coddington equations and Snell's law against wavelength. The limitation of the method is that the chromatic aberrations of the fields outside the symmetric plane

cannot be solved. However, they are of great interest for analyzing the asymmetric effects introduced by breaking the symmetry and the local influence of freeform surfaces on the chromatic aberrations.

2.5 Obscuration in off-axis mirror systems

The design of off-axis mirror systems are popular because it overcomes the central obscuration. Utilizing the correction ability of freeform with high degree-of-freedom, it is able to achieve diffraction limited image for a large clear aperture and small F-number [2-17]. It is shown in section 2.4.2 that, on the one hand, the tilting angle of the mirror influences the aberrations; on the other hand, the tilting angle cannot be too small to block the ray path. In order to embed obscuration constraint in optimization, an error function, which describes the truncation between mirrors and ray bundles, has to be introduced. There are plenty of ways to establish the error function. As early as 1987, Rodgers has put forward one method to design unobscured plane symmetric reflective systems by solving the linear inequalities between boundary rays and mirror edges [2-18]. However, this method is only valid for the off-axis mirror systems with a zigzag structure. In the folded structure, comparing the y -height between the mirror and the boundary ray is meaningless. In recent years, many more attempts have been done to minimize aberrations as well as remove obscuration in plane symmetric mirror systems automatically during optimization [1-15, 2-19, 2-20].

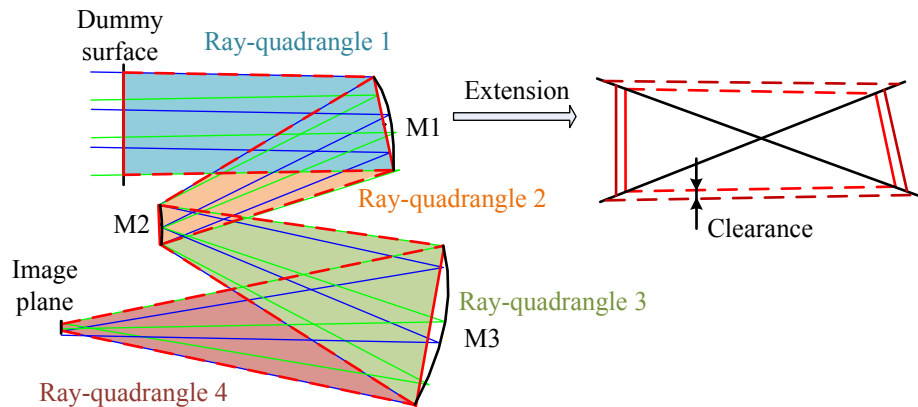


Fig. 2-11 Construction of ray-quadrangles in a three-mirror telescope.

One of the interesting methods is developed by Xu *et al.*, which considers the overlap between the ray-quadrangles. A three-mirror off-axis system is shown in Fig. 2-11 for illustration. The ray-quadrangles are built by connecting the upper and lower boundary ray intersection points on the two neighboring mirrors in the tangential plane. A dummy surface in front of the first mirror and the image plane help build the first and the last ray-quadrangles, respectively. In this way, the whole ray path in the system is divided into four ray-

quadrangles. The two sides that contain the mirrors are called “hard side”. The two ray-sides are called “soft side”. In order to cover the vertices of concave mirrors, which are outside of the ray-quadrangles (e.g., M1 and M3), and leave some space for mechanical mounting, each side of the ray-quadrangles is extended outwards by certain clearance. The four extended sides and the two diagonals of the original ray-quadrangle cut the extended area into four triangles. There are three types of error functions to describe the overlap between the ray-quadrangles.

The first two error functions consider the position relations between the two non-neighboring ray-quadrangles a and b in Fig. 2-12. If a vertex P of the ray-quadrangle b is located inside the extended area of a , the squared distance from P to the corresponding extended side is calculated. Therefore, the first error function is written as

$$Erf_{2D1}(a, b) = \sum_{j=1}^4 \begin{cases} |P_j E|^2 + \text{Bias}, & \text{if } P_j \in \Lambda \\ 0, & \text{if } P_j \notin \Lambda \end{cases} \quad (2.24)$$

where $j = 1$ to 4 is the sequence of vertices of the ray-quadrangle b , Λ is the extended area of the ray-quadrangle a , E is the cross point between the vertical line and the corresponding extended side. The positive bias ensures the clear separation between the zero and non-zero parts.

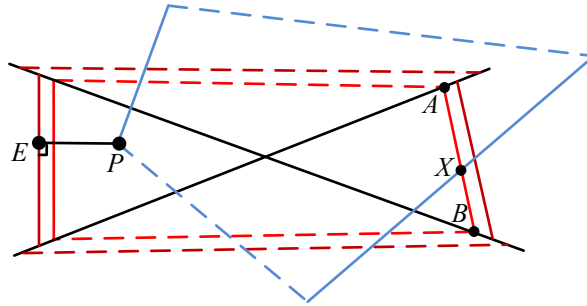


Fig. 2-12 Two situations of obscuration between the ray-quadrangles a (red) and b (blue): (1) the vertex P is inside a , (2) one side of b cross the hard side of a at the point X .

The second situation describes the cross between any side of the ray-quadrangle b and the two hard sides of the ray-quadrangle a . If the cross happens, the cross point X cuts the hard side AB into two parts $|AX|$ and $|BX|$. The second error function takes the shorter part into account, which has the form

$$Erf_{2D2}(a, b) = \sum_{k=1}^2 \sum_{j=1}^4 \begin{cases} \min(|A_k X_j|^2, |B_k X_j|^2) + \text{Bias}, & \text{if crossed} \\ 0, & \text{if uncrossed} \end{cases} \quad (2.25)$$

where $k = 1$ to 2 is the two hard sides of the ray-quadrangle a , $j = 1$ to 4 is the sequence of four sides of the ray-quadrangle b .

After checking all the combinations of non-neighboring a and b , the only unconsidered case left is the position relation between the image plane and the $(N_q-1)^{\text{th}}$ ray-quadrangle (assuming a reflective system has N_q ray-quadrangles). The first error function can be used to detect this case if a and b are substituted to $N_q - 1$ and N_q , respectively. Then, the third error function has the following form

$$Erf_{2D3}(a, b) = Erf_{2D1}(N_q - 1, N_q) = \sum_{j=3}^4 \begin{cases} |P_j E|^2 + \text{Bias}, & \text{if } P_j \in \Lambda \\ 0, & \text{if } P_j \notin \Lambda \end{cases} \quad (2.26)$$

where $j = 3$ to 4 shows that only the two vertices belonging to the image plane are checked.

The advantage of the error functions developed by Xu *et al.* is that it can be optimized together with the image quality constraints to obtain the initial system design. However, retrieving back to section 2.2, the algorithm can only deal with the reflective systems shown in Table 2-3. The auto-optimization of the more generalized reflective systems, such as Yolo telescope, cannot be realized.

2.6 Modelling of real freeform surfaces

Freeform surfaces are commonly manufactured by the ultra-precise diamond turning machine. Unlike the fabrication process of rotationally symmetric surface, which only generates concentric errors, the involvement of fast tool servo (FTS), slow tool servo (STS), and the thermal effect cause anisotropic errors and ripples in different directions [2-21, 2-22]. Fig. 2-13 shows the manufacturing surface error $\Delta z(x, y)$

$$\Delta z(x, y) = z_{\text{real}}(x, y) - z_{\text{ideal}}(x, y) \quad (2.27)$$

of a diamond-turned surface, where $z_{\text{real}}(x, y)$ is the topography of the manufactured surface measured by interferometry, $z_{\text{ideal}}(x, y)$ is the designed surface described by the polynomial functions introduced in section 2.1.3. Then, by applying polishing techniques, such as computer-controlled polishing (CCP) or magnetorheological finishing (MRF), the accuracy of the manufactured surface can be further improved.

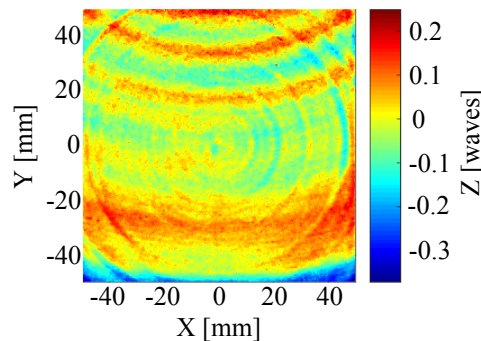


Fig. 2-13 Manufacturing surface error of a diamond-turned surface.

The manufacturing quality of a real surface can be represented by the peak-valley (PV) value, which is the difference between the largest and the smallest surface deformation values, and the RMS value of the surface errors. In order to quantify the composition of the surface errors, the power spectral density [2-23] is applied, which shows the power distribution of the surface errors in the frequency domain. Other techniques to specify the surface errors include area structure function [1-20, 2-24, 2-25] and polar RMS [1-21].

2.6.1 Power spectral density

Before calculating the 2D discrete Fourier transform (2D-DFT) of the surface errors, zero padding with a proper window such as Butterworth filter is applied to ensure the smooth change of the data to zero at the boundary, if it is not rectangular. The 2D-DFT of a N_x by N_y equal sampling data set is written as [2-22]

$$U(v_x, v_y) = \sum_{n_x=0}^{N_x-1} \sum_{n_y=0}^{N_y-1} \Delta z(x, y) e^{i2\pi\left(\frac{v_x x}{N_x} + \frac{v_y y}{N_y}\right)} \quad (2.28)$$

where v_x and v_y are the frequencies in the x - and y -directions. The 2D PSD [44]

$$\text{PSD}_{2D}(v_x, v_y) = \frac{1}{A} |U(v_x, v_y) \Delta x \Delta y|^2 = \frac{|U(v_x, v_y)|^2 D_x D_y}{N_x^2 N_y^2} \quad (2.29)$$

is the normalized energy distribution over the frequency range

$$v_x = -\frac{1}{2\Delta x} \text{ to } \frac{1}{2\Delta x}, \quad v_y = -\frac{1}{2\Delta y} \text{ to } \frac{1}{2\Delta y} \quad (2.30)$$

with the spacing in the frequency domain

$$\Delta v_x = \frac{1}{D_x}, \quad \Delta v_y = \frac{1}{D_y}. \quad (2.31)$$

From Eq. (2.29), the 2D PSD has the unit of amplitude squared per unit frequency squared. Δx and Δy are the sampling spacing in the spatial domain, D_x and D_y are the widths of the surface in the x - and y -directions, A is the area equaling to $D_x \times D_y$.

It is possible to integrate the 2D PSD along any line going through the center of the frequency domain with arbitrary azimuthal angle to get the 1D PSD. For example, the integrated 1D PSD along the y -direction is written as [2-23]

$$\text{PSD}(v) = \sum_{n_x=0}^{N_x-1} \text{PSD}_{2D}(v_x, v_y) \Delta v_x, \quad (2.32)$$

which has the unit of amplitude squared per unit frequency. The RMS surface error σ_{rms} is the square root of the area under the 1D PSD curve, and can be written as [2-23]

$$\sigma_{\text{rms}} = \sqrt{\sum \text{PSD}(\nu) \Delta \nu} . \quad (2.33)$$

The typical 1D PSD curve of the surface errors is sketched in Fig. 2-14 as a log-log diagram. The curve is mainly linear but with nonlinear deviation on the left end. The slope of the line depends on the fabrication machine. The whole spectral range is divided into LSF, MSF, and HSF ranges.

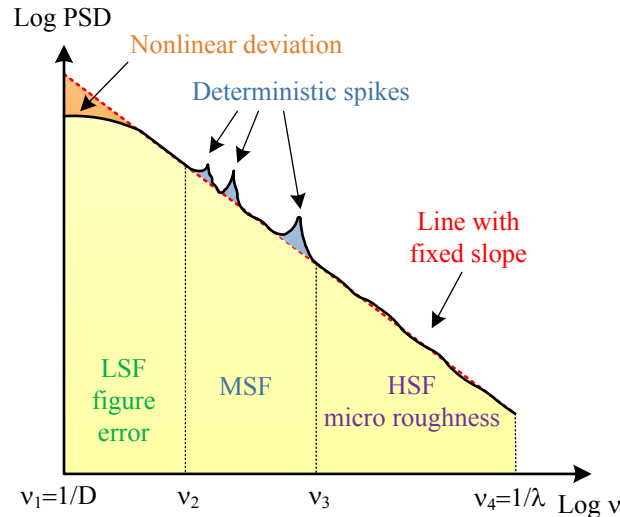


Fig. 2-14 1D integrated PSD in the log-log diagram [2-26, 2-27].

The LSF errors are figure errors, for example, radial and asymmetric irregularities, which act like classical aberrations. The LSF errors are fully deterministic and result in a decrease in resolution. The left boundary frequency depends on the size of the surface.

The HSF errors are the micro roughness on the surface and purely statistical. They cause the drop of contrast of the system. Theoretically, the maximum frequency equals one over the measuring wavelength. However, the maximum frequency that can be resolved is limited by the spatial resolution shown in Eq. (2.30).

The MSF errors consist of both random error, which is the reason for the linear decay of the PSD, and the deterministic spikes, which is due to the dynamics of the diamond turning machine and contributions of higher order Zernike deformations. Therefore, the influence of the MSF errors on the system performance comes from the mixture of both statistical and deterministic errors.

2.6.2 Description of low spatial frequency errors

As mentioned, the LSF errors provoke classical aberrations. Therefore, the polynomial functions have the capacity to describe this type of errors. Especially, Zernike polynomials and Q-polynomials are preferred because of their orthogonality.

2 Theoretical Introduction

The direct relation between the surface RMS value and the Zernike Fringe coefficients $a_{n,m}$ in Eq. (2.7) is that

$$\sigma_{\text{rms}} = \sqrt{\sum_{n=0}^N \frac{a_{n,0}^2}{n+1} + \frac{1}{2} \sum_{n=0}^N \sum_{m=-n}^n \frac{a_{n,m}^2}{n+1}}. \quad (2.34)$$

Assuming a set of surfaces, each of them is described by a single Zernike Fringe term with a PV value of 1mm. If the PSD of each surface is plotted together, it is interesting to see from Fig. 2-15 that the PSDs are decreasing on the same envelope line in red. The peak intensities of the curves (see yellow dots and orange line) are located slightly different.

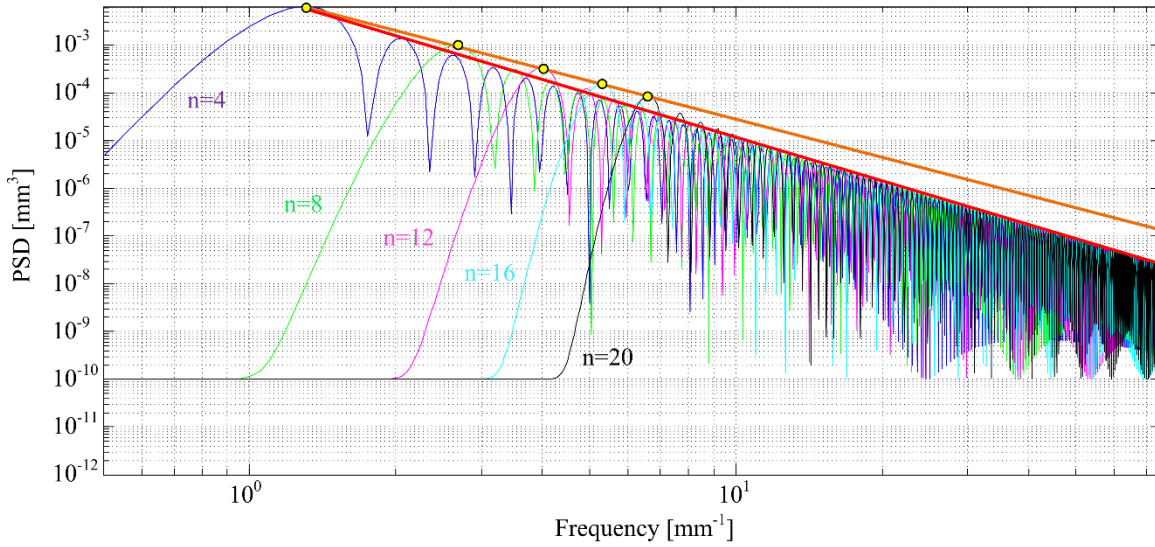


Fig. 2-15 PSD of the Zernike fringe polynomial terms with radial orders $n = 4, 8, 12, 16$ and 20 [2-26].

Alternatively, the LSF errors can be represented by the lateral shifted radial basis functions (RBFs)

$$G(x, y) = \sum_m a_m e^{-\frac{(x-x_m)^2 + (y-y_m)^2}{w^2}} \quad (2.35)$$

where (x_m, y_m) is the sampling grid for fitting the errors. Usually, the Cartesian, polar, hexagonal, Fibonacci, and statistical grids are used. w is the shape factor that decides the width of the Gaussian functions. Therefore, the number of sampling points, the choice of sampling grid and the width determine together the accuracy of the fitting. Fig. 2-16 shows the fitting of a given profile in one dimension. It is also possible to use an unequal sampling grid to improve the accuracy.

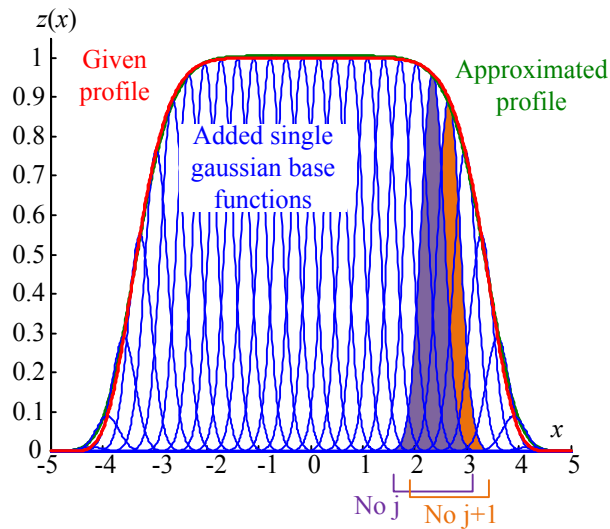


Fig. 2-16 1D profile fitting with equally sampled RBFs [2-26].

2.6.3 Description of deterministic ripples

The MSF ripples have a periodic structure. Therefore, the frequencies of the ripples are fixed for a specific machine. According to the 1D description in [2-28] and extension into 2D, the mathematical functions of different types of the ripples are listed in Table 2-5.

For all the groove functions, they have the PV height of h and period of T . The radius is given by

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2} \tag{2.36}$$

where (x_c, y_c) is the center of the groove structure, which shows that the center of the ripple is not necessarily coinciding with the center of the freeform surface.

| Types | Functions | Shapes |
|----------------------------|---|--|
| Rectangular | $\frac{h}{2} \left\{ \text{Sgn} \left[\sin \left(\frac{2\pi r}{T} \right) \right] + 1 \right\}$ | $h \left[\frac{\text{Square wave}}{T} \right]$ |
| Triangular | $\frac{h}{2} \text{Sin}^{-1} \left[\text{Sin} \left(\frac{2\pi r}{T} \right) \right]$ | $h \left[\frac{\text{Sawtooth wave}}{T} \right]$ |
| Sinusoidal | $\frac{h}{2} \left[\text{Cos} \left(\frac{2\pi r}{T} \right) + 1 \right]$ | $h \left[\frac{\text{Cosine wave}}{T} \right]$ |
| Piecewise-parabolic | $h \left\{ \text{sin}^{-1} \left[\text{Sin} \left(\frac{\pi r}{T} \right) \right] \right\}^2$ | $h \left[\frac{\text{Parabolic wave}}{T} \right]$ |

Table 2-5 Mathematical functions of different types of MSF ripples [2-28].

2.6.4 Description of random errors

The random errors in MSF and HSF range follow the statistical random model with known amplitudes according to the linear PSD and unknown random phases in the frequency

domain. Assuming the random errors have the RMS value of $\sigma_{\text{rms, rand}}$. The 2D random phase ϕ is generated by

$$\phi = 2\pi \cdot \text{Rand} \quad (2.37)$$

where Rand is the N_x by N_y matrix with random numbers between 0 and 1 in the continuous uniform distribution. There are two ways to generate random errors [2-29].

The direct method is related to the slope factor k , which determines the linear PSD in the semi-logarithmic plot. The amplitudes in the 2D grid can be calculated by

$$A_m = 10^{\frac{k}{2}} \sqrt{v_x^2 + v_y^2}. \quad (2.38)$$

The 2D Fourier components of the random errors can be written as

$$U = \frac{A_m}{\sqrt{\sum A_m}} e^{i\phi}. \quad (2.39)$$

Then, applying the inverse fast Fourier transform (iFFT) and taking only the real part, the surface is reconstructed in the spatial domain as

$$z'_{\text{rand}} = \text{real}(\text{iFFT}(U)) \quad (2.40)$$

with the RMS value $\sigma'_{\text{rms, rand}}$. The random errors are scaled by the correct RMS value as

$$z_{\text{rand}} = \frac{\sigma_{\text{rms, rand}}}{\sigma'_{\text{rms, rand}}} z'_{\text{rand}}. \quad (2.41)$$

The alternative way to generate random errors is based on convolution. Generating a random array scaled by $\sigma_{\text{rms, rand}}$ as

$$A_1 = \sigma_{\text{rms, rand}}(2 \cdot \text{Rand} - 1). \quad (2.42)$$

Defining another array in the Gaussian form as

$$A_2 = \frac{1}{L_c \sqrt{\pi}} e^{-\frac{x^2 + y^2}{2L_c^2}} \quad (2.43)$$

with the correlation length L_c , which is related to the slope of the PSD curve. Then, the random errors are the convolution between the two arrays as

$$z'_{\text{rand}} = A_1 * A_2. \quad (2.44)$$

Again, the random errors have to be scaled by the correct RMS value according to Eq. (2.41). The PSD of the random error obtained by this method also has the Gaussian shape, which means that the left end of the curve is bended. The random errors generated by the above two methods are compared in Fig. 2-17.

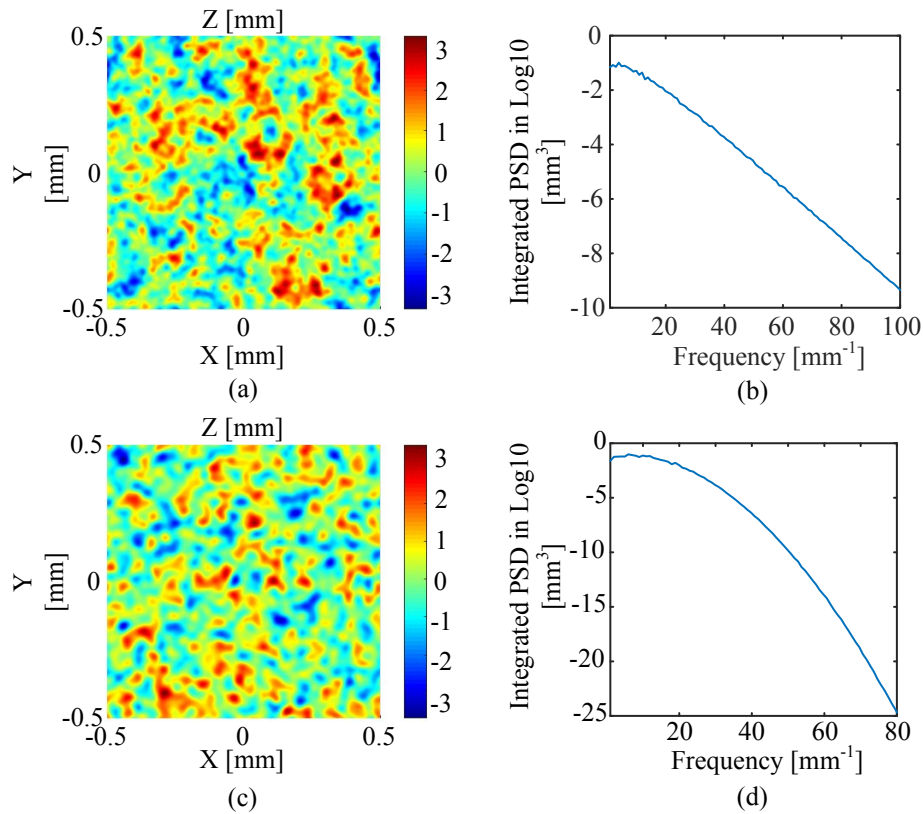


Fig. 2-17 Random surface errors (a) and its 1D integrated PSD (b) generated by the direct method. Random surface errors (c) and its 1D integrated PSD (d) generated by the convolution method.

2.6.5 System performance criteria

Different types of surface errors have their own influence on the system resolution and contrast. It is demanding to evaluate the degradation of the system performance after manufacturing.

The simplest way is to investigate the spot diagram. The shape of the spot is determined by the aberrations. For example, spherical aberration, coma, and astigmatism result in circular, water drop-like, elliptical spots, respectively. However, when the system is influenced under LSF errors with multiple lower order aberrations, it is nearly impossible to identify the specific aberration from the shape of the spot. A more feasible way is to consider the enlargement of the RMS spot radius, which computes the RMS value of the ray intersection distances to the chief ray or centroid ray intersection point in the image plane. Another difficulty is that the redistribution of the energy density in the image plane as the consequence of MSF ripples and HSF errors cannot be easily resolved from the spot diagram.

The point spread function (PSF) shows the energy distribution in the image plane from a single point source. Precisely, the PSF is computed according to the Huygens principle that superimposes the Huygens wavelets in the image plane. If the far-field approximation is

fulfilled, the PSF can be computed by the Fourier transform of the complex function in the exit pupil, which is much faster.

The influence of surface errors in different frequency regions on the PSF is sketched in Fig. 2-18. The PSF of the nominal system is indicated by the dashed curve. The classical aberrations in LSF errors decrease the intensity and broadens the central peak. The MSF errors induce small-angle scatter, thus broadening the central core and reducing the resolution and create new side lobes far away. The HSF errors produce wide-angle scatter and reduce the contrast. [2-30]

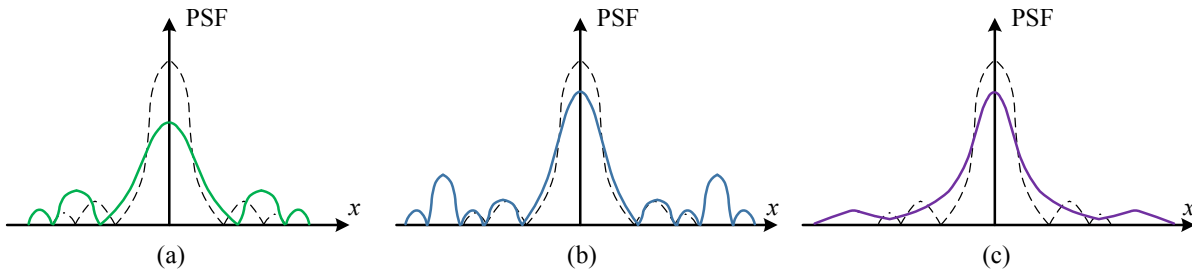


Fig. 2-18 Change of the nominal PSF (dashed curve) due to (a) LSF, (b) MSF, and (c) HSF errors [2-30].

The PSF describes the image quality in the spatial domain. The modulation transfer function (MTF), which is the amplitude of the Fourier transform of the PSF, describes the response of the optical system in the frequency domain. The MSF ripples can be viewed as the overlay of a grating on the optical surface. It causes diffraction effects. As a result, the non-nominal diffractive orders have higher intensity in PSF, and the MTF curve shows the waviness [2-31, 2-32]. The periodic MTF is analytically derived by Tamkin *et al.* as [2-31]

$$\text{MTF}_{\text{ripple}}(\nu) = \left| \text{tri} \left(\frac{\lambda l' \nu}{D_{\text{Exp}}} \right) \cdot \sum_{m=-\infty}^{\infty} J_m^2 \left(\frac{2\pi}{\lambda} h \right) \cdot \cos \left(2\pi \frac{\nu m \lambda l'}{T} \right) \right| \quad (2.45)$$

where l' is the image distance, D_{Exp} is the diameter of the exit pupil, m is the diffractive order. It shows that the drop of MTF is related to h , which is the PV of the ripples.

The drop of MTF due to the wide-angle scatter of the random errors was investigated by Harvey. The surface transfer function (STF) is defined as [2-30, 2-33]

$$\text{STF} = e^{-(4\pi\sigma_{\text{rms}})^2} \frac{1 - \text{ACV}}{\sigma_{\text{rms}}^2} \quad (2.46)$$

in the frequency domain, where ACV is the surface autocovariance function. The MTF effected by surface scatter is the product of STF and the ideal MTF [2-30]

$$\text{MTF}_{\text{random}} = \text{STF} \cdot \text{MTF}_{\text{ideal}} \quad (2.47)$$

2.7 Open questions

The development of symmetry-free optical systems promotes the generalization of the system description, optimization constraints, and tolerancing. Many attempts have been made as introduced in this chapter, but the work has not been accomplished yet. Table 2-6 lists the current status of each research direction concerning imaging equation, aberration theory, and obscuration with regard to the degradation of system symmetry.

| System types | | Issues | | | |
|--------------------------|---------------|------------------------|------------------------|-----------------|---------------|
| | | Rotationally symmetric | Double-plane symmetric | Plane symmetric | Non-symmetric |
| Surface imaging equation | | | | | |
| Aberration theory | Monochromatic | | | | |
| | Chromatic | | | | |
| Obscuration | | | | | |
| Tolerancing | | | | | |

Table 2-6 Current status of the research for different system types. Green, yellow, and red indicate that the issue is fully, partly, or not solved. The issue of obscuration is unavoidable in rotational and double-plane symmetric optical systems with reflective surface.

It can be clearly seen that the surface imaging equations and monochromatic aberrations are successfully described for all types of systems. The remaining issues are

1. Obscuration elimination in non-rotationally symmetric mirror systems

Symmetry-free Yolo telescope could also suffer from obscuration during the design process. Nevertheless, no force focuses on this point at the moment. Chapter 3 extends Xu's method and puts forward an automatic optimization algorithm, which is able to remove the obscuration in an off-axis mirror systems without symmetry.

2. Systematical description of the generalized chromatic aberrations in non-rotationally symmetric optical system

Although the chromatic aberrations are analytically derived by Seidel for rotationally symmetric systems, and partly investigated by Cao *et al.* for the fields in the symmetric plane of double-plane and plane symmetric systems, the systematic description of the generalized chromatic aberrations for systems with arbitrary symmetry is still missing. Chapter 4 puts an insight on this topic, gives a clear definition of the chromatic aberrations in the general form, and proposes both ray- and wavefront-based methods to solve the chromatic aberrations surface-by-surface for the whole field of view and arbitrary surface shape.

Open questions also arise when moving from the design phase to the production (numbered following the above second point):

3. The boundary frequencies between different surface error types

In practice, the LSF errors are in the range between $\frac{1}{D}$ and $\frac{10}{D}$, MSF errors cover the region between $\frac{10}{D}$ and $\frac{50}{D}$, the higher frequency region belongs to HSF errors according to experience. This is dependent on the absolute size of the surface diameter. However, the method of finding out the concrete boundary frequencies for an individual machine is unclear.

4. The comprehensive model for all types of surface errors

There are several models to describe the manufacturing errors of freeform surfaces and several matrices to evaluate the impacts of each type of error on the system performance. However, tolerancing requires the overlay of all types of errors simultaneously. The investigation of the image quality with simulated freeform surface errors has not been seen in the literature.

These two questions will be answered in chapter 5.

3 General Solution for Obscuration Constraints

The method described in section 2.5 can only eliminate the obscuration in plane-symmetric optical systems. To deal with the more general systems in the outermost ring of Fig. 2-1, the method has to be extended by constructing the error function, which describes the degree of obscuration in 3D. Then, two examples are provided. In the first example, an unobscured three-mirror telescope layout is produced by optimizing the error function to be zero within the selected ranges of the variables. In the second example, the obscuration elimination is combined with astigmatism correction to obtain a proper initial Yolo telescope setup.

3.1 Construction of error functions

In off-axis plane-symmetric optical systems, the optical elements rotate around one axis. Therefore, the globally symmetric plane (tangential plane) is always used to cut the optical elements and find the upper and lower ray intersection points on the surfaces, which construct ray-quadrangles to detect the obscuration. The obscuration is not occurring in the sagittal plane. In the 3D case, the mirrors in the optical systems are freely rotated against local x - and y - axes. The advantages of the symmetry-free optical system are that, on the one hand, more degrees of freedom are introduced to minimize aberrations; on the other hand, a more compact system can be obtained. However, the tangential planes behind each surface are no longer lying on a common plane. This can be seen from the example of a simple 3D two-mirror Yolo telescope as shown in Fig. 3-1. The tangential and sagittal rays are plotted in red and green, respectively. Because of the 3D rotation of the mirrors, the neighboring ray-quadrangles shadowed in blue and yellow are now belonging to different planes, the algorithm introduced in section 2.5 that is limited to judging the position relation between

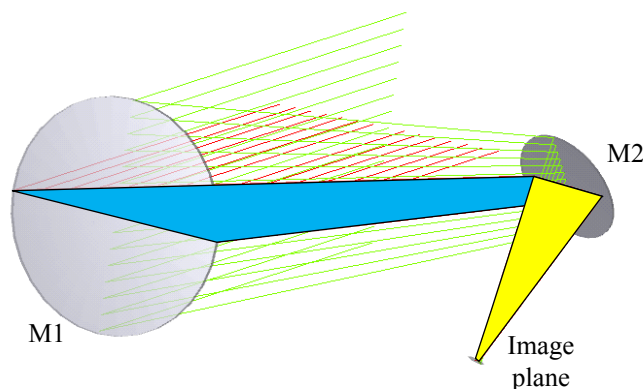


Fig. 3-1 3D Yolo telescope with M1 and M2 rotate freely in 3D space. The propagation of the tangential and sagittal ray bundles are drawn in red and green respectively. As a result, the two neighboring ray-quadrangles (blue and yellow) are no longer in the same plane.

the two ray-quadrangles in the 2D plane cannot be directly applied here. An extension of the obscuration elimination algorithm to 3D system is necessary.

The goal is to establish an error function that describes all the possible truncation of light in a symmetry-free optical system. If the obscuration happens, at least a mirror blocks the ray bundle between two adjacent surfaces. It means that, although an optical system contains many surfaces, only three surfaces (two neighboring surfaces and an additional surface) are discussed at each time. The two adjacent surfaces and the outside boundary of the ray bundle in between form an irregular volume, which is unable to be described by a rigorous mathematical function. The easiest way is to derive the 3D issue to 2D. This can be realized by the following steps as shown in Fig. 3-2:

1. The surfaces are simplified as circles by their circular boundaries. The semi-diameter of the circle is determined by the boundary ray intersecting point on it.
2. It is well-known that three non-collinear points determine a plane. For the two adjacent circles ($a, a+1$) and another circle (b), a local common plane can be found which goes through the centers of the three circles.
3. The local common plane cuts the circles into three lines. The two lines belonging to the two neighboring surfaces construct a quadrangle, which describes the range of the ray bundle.
4. The four sides of the quadrangle are extended for certain clearance to include the vertices of the mirrors, which may fall outside of the quadrangle (for example, the case of circle $a+1$), and leave some space for mounting. The two mirror sides plot in solid lines are numbered as the first and the third sides. The two ray sides plot in dashed lines are numbered as the second and the fourth sides.

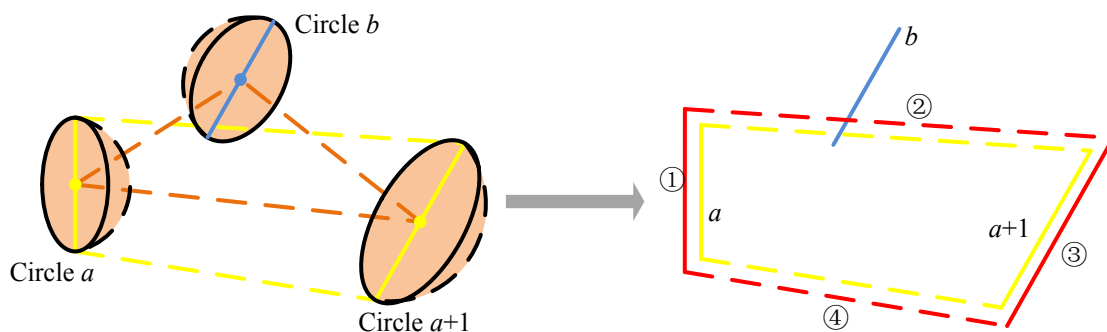


Fig. 3-2 Process of deriving the surface and ray bundle relation from 3D to 2D. The dashed black curves show the boundary of the mirrors. The numbers denote the sequence of the sides.

In this way, the position-relations between a ray bundle and a surface are deduced to be the position-relations between a quadrangle and a line in the same plane. All the possible cases are plotted in Fig. 3-3. For the cases (a) to (f), either the line intersects the quadrangle, or at least one endpoint of the line is located inside the quadrangle, which shows the overlap of the mirror to the ray bundle. To detect whether an endpoint P of the line is inside the quadrangle, the error function Eq. (2.24) introduced in section 2.5 can be used here. The two diagonals divide the quadrangle into four triangles. If P is inside any of the triangles, the distance $|PE|$ from the point P to the corresponding side of the quadrangle is calculated. Similarly, the first error function in the 3D system is written as

$$Erf_{3D1}(a, b) = \sum_{j=1}^2 \begin{cases} |P_j E|^2 + \text{Bias}, & \text{if } P_j \in \Lambda \\ 0, & \text{if } P_j \notin \Lambda \end{cases} \quad (3.1)$$

where a, b are the indices of the quadrangle and the additional surface, j represents the endpoint on the line b , Λ is the area of the quadrangle.

In Fig. 3-3 (a), (b), (d) - (f), the line crosses at least one side of the quadrangle. For example, the line crosses the side AB at the point and divides the side into two parts AX and BX . The case can be described by the error function Eq. (2.25). When discussing the relation between two ray-quadrangles in 2D, the cross between the object hard side and subject soft side is not considered to avoid redundancy because the object and the subject ray-quadrangles will exchange with each other. However, the crosses between the line and all the four sides of the quadrangle are considered in 3D. Then we get the second error function in the 3D system, which is the square of the shorter part length, as

$$Erf_{3D2}(a, b) = \sum_{k=1}^4 \begin{cases} \min(|A_k X|^2, |B_k X|^2) + \text{Bias}, & \text{if crossed} \\ 0, & \text{if uncrossed} \end{cases} \quad (3.2)$$

where k denotes the number of sides of a quadrangle.

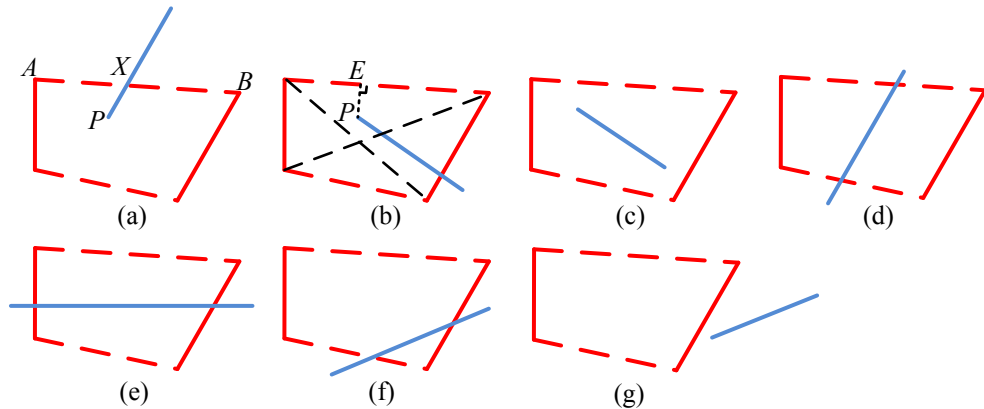


Fig. 3-3 Enumeration of the position relation between a ray-quadrangle and a line.

3 General Solution for Obscuration Constraints

Based on the above discussion, the situation that the line blocks the quadrangle in 2D is fully described. In the last case shown in Fig. 3-3 (g), the line is outside the quadrangle. However, it does not mean that there must be no truncation in 3D. The reason is that, in the process from 3D to 2D, information in one dimension is lost. The truncation in 2D proves the truncation in 3D. The reverse does not hold true. The possible situations in which the surface (/circle) b blocks the ray path in 3D but the line b is outside the quadrangle in 2D are drawn in Fig. 3-4. In Fig. 3-4 (a), the surface b intersects one of the surfaces which form the volume; in Fig. 3-4 (b), the surface b truncates the ray bundle only.

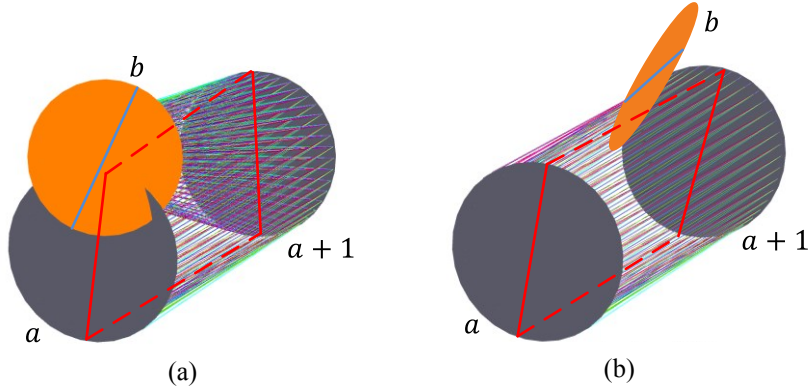


Fig. 3-4 The cases that the surface blocks the ray bundle in 3D but there is no truncation between line and quadrangle in 2D. The clearance is set to be zero in the drawing.

In order to avoid tremendous calculation when the line is outside of the quadrangle, the concept of “absolute non-intersection region” where the single surface has absolutely no influence on the ray path is defined. When a circle rotates around its center, it is inside a sphere with the radius equal to the circle's radius plus the clearance. If the center distance between the two spheres is longer than the sum of their radii, the spheres are away from each other so that the intersection between the two circles is avoided. Therefore, the first condition that the circle b is located inside the “absolute non-intersection region” can be written as

$$r_{a,b} > r_a + r_b, \quad r_{a+1,b} > r_{a+1} + r_b \quad (3.3)$$

where r_a, r_{a+1}, r_b are the radii of spheres of the circles denoted by the subscript, $r_{a,b}, r_{a+1,b}$ are the distances between the centers of spheres denoted by the subscript. The line outside of the quadrangle means that the center of the circle b is outside of the quadrangle. If the distances between the center of circle b and the two soft sides of the quadrangle respectively are longer than r_b , the circle b cannot touch the ray bundle. Accordingly, the second condition can be written as

$$t_{k,b} > r_b \quad (3.4)$$

where $k = 2,4$ represent the two soft sides of the quadrangle. If both Eq. (3.3) and (3.4) are fulfilled at the same time, as shown in Fig. 3-5, the circle b is located inside the “absolute

non-intersection region” so that no further check about obscuration is needed. Otherwise, the two cases shown in Fig. 3-4 have to be discussed individually.

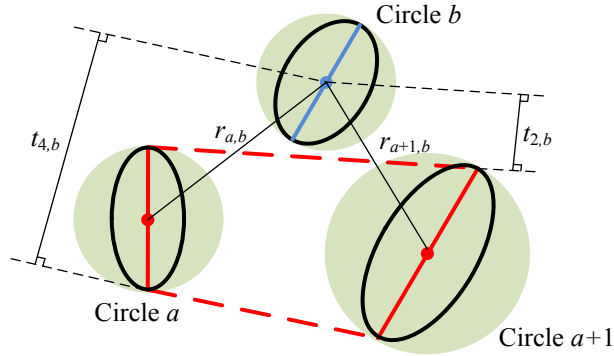


Fig. 3-5 Circle b is in the “absolute non-intersection region”. The clearance is set to be zero in the drawing.

The third error function is used to describe the possible intersection between the two circles. In Fig. 3-6 (a), two circles are inside two non-parallel planes. The line of intersection between the two planes is calculated. If there is a common intersection part among the line of intersection and both circles, the length of the common part is calculated, and the 3rd error function is written as

$$Erf_{3D3}(a, b) = \sum_{i=a}^{a+1} \begin{cases} |P_{i,1}P_{i,2}|^2 + \text{Bias}, & \text{if crossed} \\ 0, & \text{if uncrossed} \end{cases} \quad (3.5)$$

where $P_{i,1}$ and $P_{i,2}$ are two endpoints of the common intersection part. The possible intersections between the circle b and circle $a / a+1$ are checked respectively.

The case in Fig. 3-4 (b) that the mirror only blocks the ray path is the most complicated one. The line divides the circle b into two parts. The half edge of the circle close to the ray bundle is sampled from an endpoint P_1 to another point P_2 according to the azimuthal angle to P_1 as shown in Fig. 3-6 (b). Each point P_θ constructs a local common plane with the centers of the two successive circles a and $a+1$. This new local common plane cuts the two successive circles into two new lines which form a new quadrangle. If the point P_θ is inside the new quadrangle with extended clearance, based on the principle introduced for $Erf_{3D1}(a, b)$, the squared distance from the point P_θ to the corresponding side is calculated and then compared with the squared distance that computed with the next point on the edge of circle to find the maximum value. The 4th error function is written as

$$Erf_{3D4}(a, b) = \max_{\theta} \begin{cases} |P_\theta E|^2 + \text{Bias}, & \text{if } P_\theta \in \Lambda \\ 0, & \text{if } P_\theta \notin \Lambda \end{cases} \quad (3.6)$$

where $\theta = 0^\circ$ to 180° is the azimuthal angle between the starting point P_1 and P_θ .

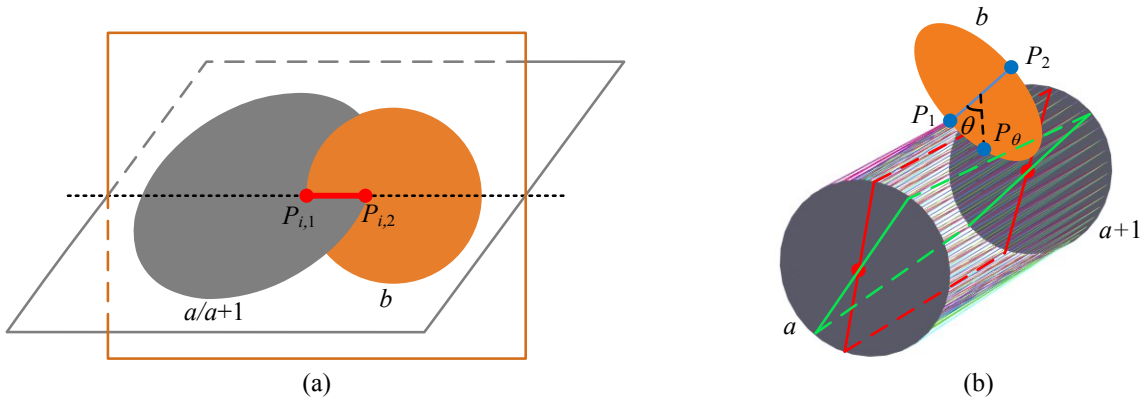


Fig. 3-6 Ways to check the intersection between mirrors in (a) and the intersection between a ray bundle and a mirror in (b).

Above all, four error functions are described to determine the possible obscuration cases between the ray bundle and an arbitrary surface in 3D. They are added by multiplying their own weighting factors to obtain an overall obscuration error function as

$$Erf_{3D}(a, b) = \omega_1 \cdot Erf_{3D1}(a, b) + \omega_2 \cdot Erf_{3D2}(a, b) + \omega_3 \cdot Erf_{3D3}(a, b) + \omega_4 \cdot Erf_{3D4}(a, b) \quad (3.7)$$

where ω_1 to ω_4 are weighting factors to adjust the ratio among the sub-error functions. They have to be not smaller than 1 to ensure a significant positive bias inside each sub-error function so that a clear gap between the obscured and unobscured cases is generated. The flowchart in Fig. 3-7 illustrates the steps to compute $Erf_{3D}(a, b)$. For the given values of a and b , three surfaces (a , $a+1$ and b) are selected. At first, the three surfaces are deduced to 2D to compute $Erf_{3D1}(a, b)$ and $Erf_{3D2}(a, b)$. If the sum of the first two error functions is non-zero, obscuration occurs, $Erf_{3D3}(a, b)$ and $Erf_{3D4}(a, b)$ are both zero. If the sum is zero, then the location of the surface b is checked. If it is inside the absolute non-intersection region, $Erf_{3D}(a, b)$ is directly zero. If not, $Erf_{3D3}(a, b)$ and $Erf_{3D4}(a, b)$ are further calculated to get $Erf_{3D}(a, b)$.

3.2 Workflow for optimization

This section discusses how the obscuration is computed in a complete optical system and how the optimization works to find the unobscured system structure. An optical system contains N optical surfaces. In addition, both object plane and image plane are used to construct the ray bundle in the object and image space. If the incoming beam is parallel, a dummy surface is inserted in the object space to build an artificial object plane. If there exists an intermediate image plane, the intermediate image plane is counted as surface as well. Therefore, there are in total S ($S \geq N+2$) surfaces in an optical system. To evaluate the obscuration situation in an optical system, the error functions of all the combinations of the

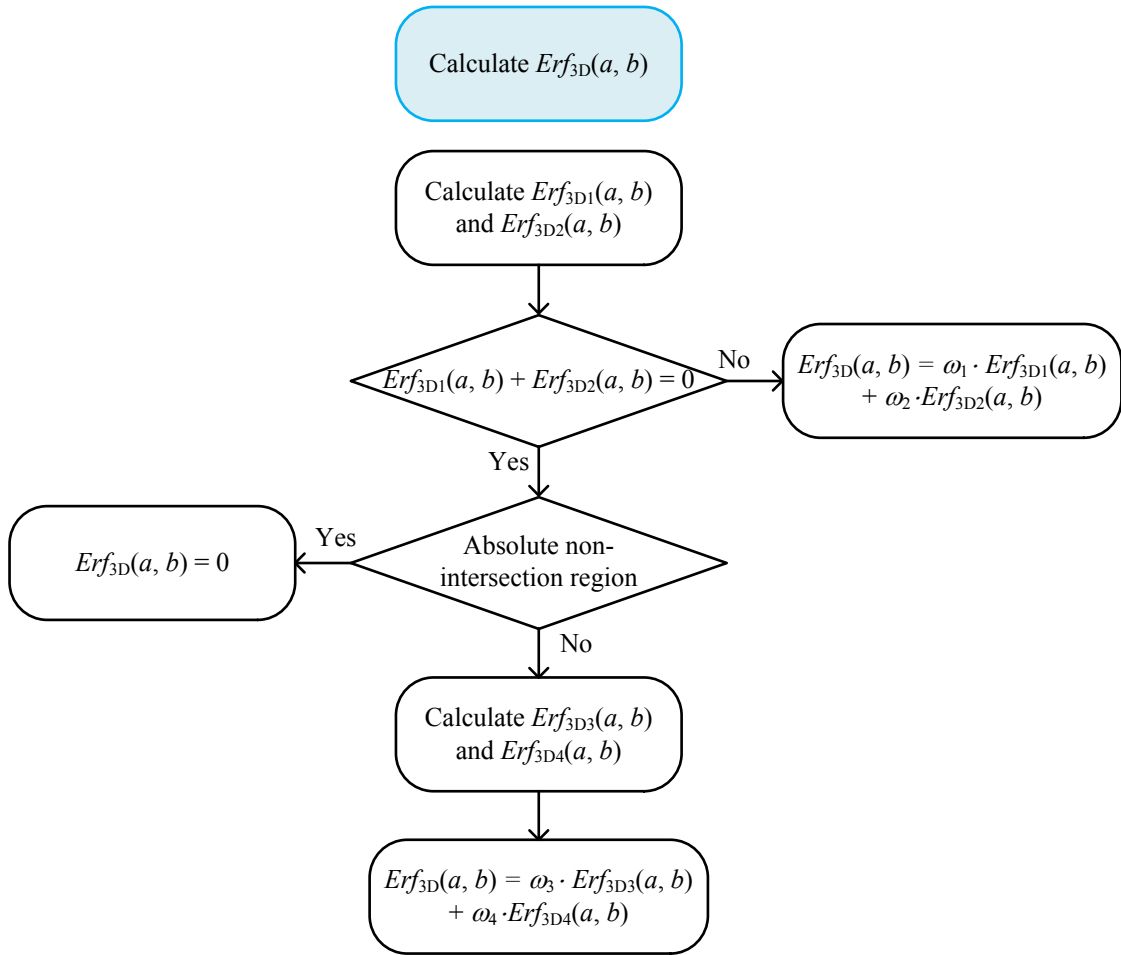


Fig. 3-7 Steps to calculate $Erf_{3D}(a, b)$.

ray bundle between surface a and $a+1$ and additional single surface b are calculated and summed up. The selection rule is

$$1 \leq a \leq S - 1, \quad 1 \leq b \leq S, \quad b \neq a \text{ and } a + 1 \quad (3.8)$$

The automatic optimization to find out the unobscured system structure is performed in MATLAB. The workflow is shown in Fig. 3-8. Before starting optimization, the variables and their ranges, the bias, and the clearance are fixed. The variables are either tilt or decenter of the optical surfaces. The goal is to find out the system structure in which the obscuration error function goes to zero. Since the obscuration error function is not monotonically decreasing, global optimization algorithms such as annealing algorithm have to be chosen to go over local minimums and find the global minimum. In each iteration (orange module), MATLAB generates new variables and inputs these variables to OpticStudio. OpticStudio does the ray-trace and outputs the global coordinates, semi-diameters of the surfaces in the current system back to MATLAB. Then the obscuration error function Erf_{3D} is calculated with the detailed steps shown in the green module. The selection rule runs to ensure that two adjacent surfaces and an additional surface are selected every time. $Erf_{3D}(a, b)$ is calculated according to Fig. 3-7 and added to the total Erf_{3D} to get the overall obscuration in the whole

system. The value of Erf_{3D} is checked whether it fulfills the requirement. If the error function goes to zero, the system under current circumstance is a good starting point for further optimization or fine-tuning.

If an additional requirement on the image quality is asked for, the obscuration error function can be combined with the image quality error function (for example, RMS spot radius, MTF, specific Zernike coefficients, etc.). The surface shape parameters, like radius, conic constant, can also be set as variables if necessary. In this case, the optimization goal is to get minimal error function value within the defined number of iterations because the error function cannot go to zero exactly because the image cannot be perfect. This will be shown in the second example.

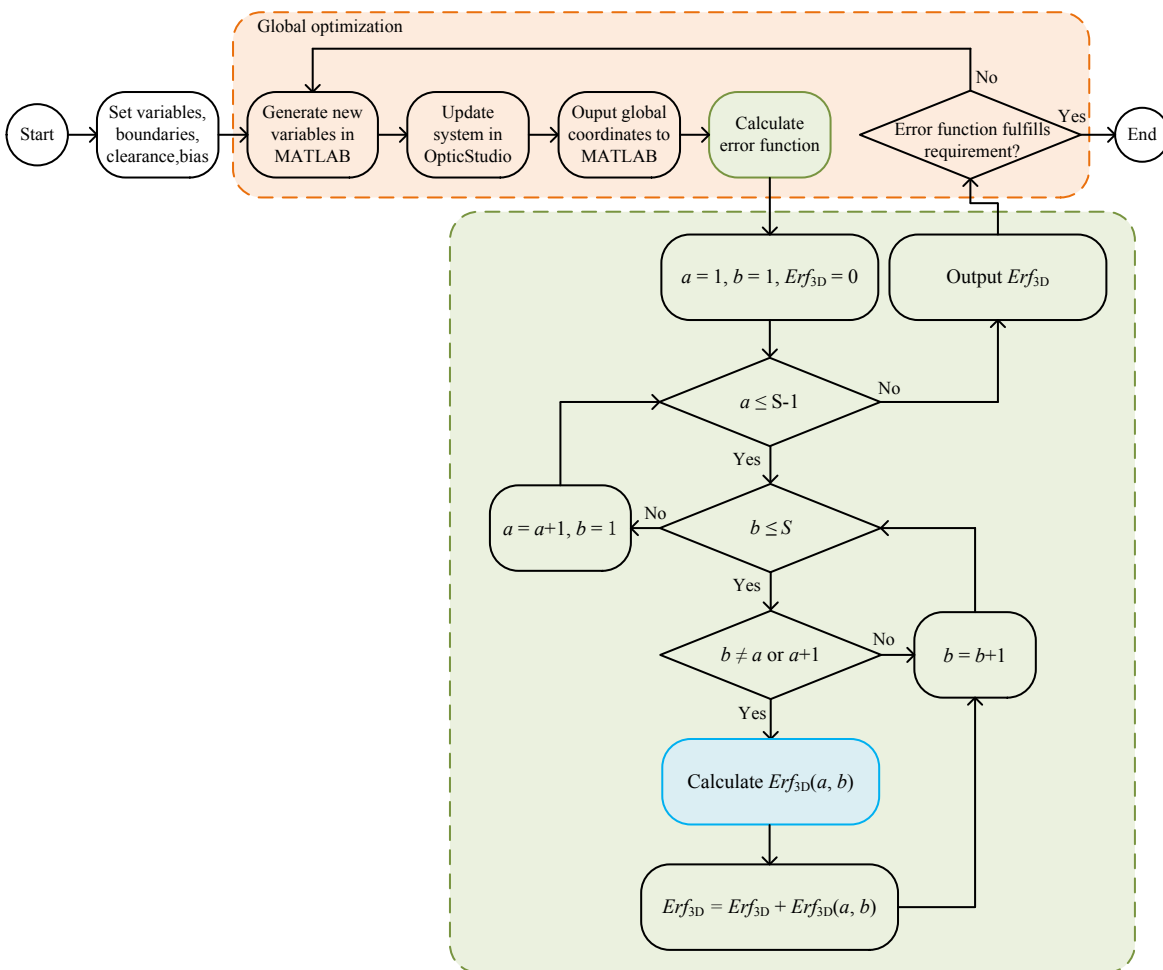


Fig. 3-8 Workflow of the automatic obscuration elimination process.

3.3 Obscuration elimination of three-mirror Yolo telescope

In the first example, the tilting angles of the mirrors in a three-mirror Yolo telescope are optimized to find out the unobscured structure. The on-axis layout of the original system is shown in Fig. 3-9 (a). The radii of curvature of the three mirrors and the image distance have

already been optimized locally to reach the best image quality for the requirements F-number 2.25 and field of view $1^\circ \times 1^\circ$. It can be seen that both M1 and image plane are fallen inside the ray bundle, M2 blocks the incoming beam totally. The aim is to optimize the tilting angles of the mirrors by the obscuration elimination algorithm to get the unobscured structure in 3D. Image quality is of no concern in the optimization process. It is assumed that the tilt of the mirror does not influence the focal length of the system.

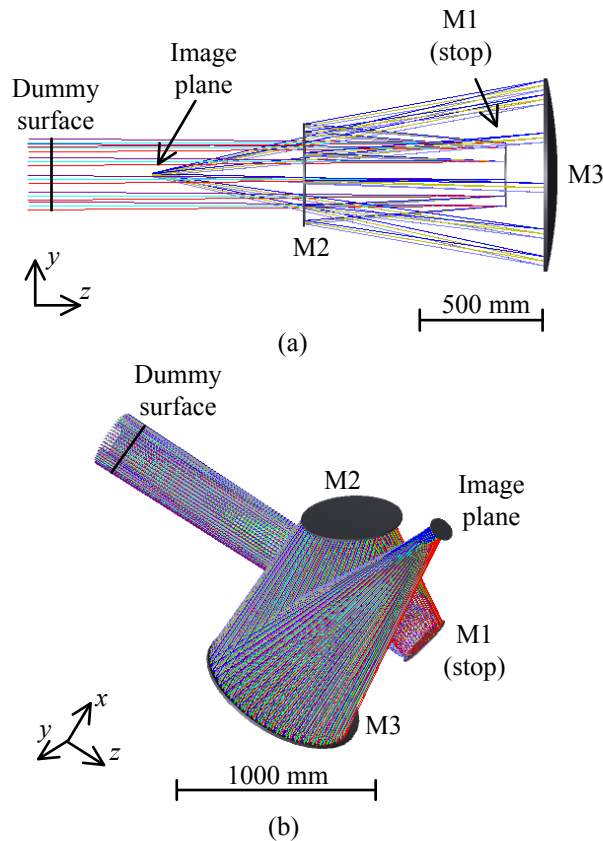


Fig. 3-9 (a) Original three-mirror telescope with all surfaces aligned on the common optical axis.
(b) The unobscured 3D system structure after optimization.

Firstly, the variables and their ranges are defined as listed in Table 3-1. The purpose of setting the ranges of the variables is to control the size of the system after optimization. The clearance is set to be 5mm, the bias is 5mm^2 , the weighting factors are all 1. The maximal iteration number is set as 100. The convergence criterion is that either the error function goes to zero or the total iteration runs out. A dummy surface is inserted in the object space to build the first quadrangle with M1 since the object is in infinity. Then the automatic optimization starts. In the first iteration when all the vertices of the mirrors are on the common axis, the plane to cut the mirrors is the tangential plane of the rotationally symmetric system. Fig. 3-10 monitors the change of error function during optimization. It can be seen that the error function drops to zero after four iterations. The tilting angles of the mirrors after optimization can be checked in the last column of Table 3-1. The layout of the optimized system is shown in Fig. 3-9 (b). It can be seen that the obscuration is removed, and the mirrors are folded to

3 General Solution for Obscuration Constraints

reduce the size of the system. The system can be utilized as an initial system for further optimization with freeform surfaces.

| Variables | Initial values | Lower boundaries | Upper boundaries | Final values |
|---------------|----------------|------------------|------------------|--------------|
| M1 tilt X (°) | 0 | 0 | 20 | 19.9331 |
| M1 tilt Y (°) | 0 | -15 | 0 | -14.6976 |
| M2 tilt Y (°) | 0 | -30 | 0 | -25.0453 |
| M3 tilt X (°) | 0 | 0 | 25 | 16.0118 |
| M3 tilt Y (°) | 0 | -20 | 0 | -1.7787 |

Table 3-1 Initial values, optimization boundaries and final values after optimization of the variables for the three-mirror Yolo telescope.

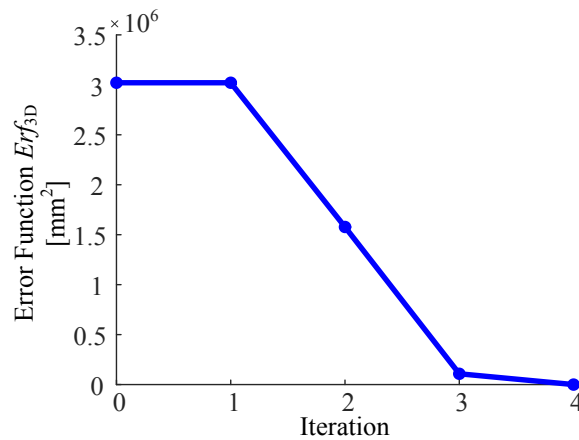


Fig. 3-10 Evolution of Erf_{3D} during optimization.

3.4 Two-mirror Yolo telescope design without astigmatism

The second example is the initial system design of a Yolo telescope with two concave mirrors. Both obscuration and astigmatism in the system are corrected by optimization to show the advantage of utilizing the 3D structure as well as the combination of the obscuration error function and the image quality error function.

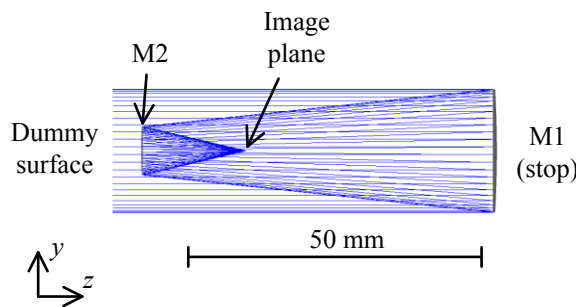


Fig. 3-11 Original two-mirror telescope with all surfaces aligned on the common optical axis.

The layout of the original system is shown in Fig. 3-11. It has only one on-axis field with the collimated incoming beam. The two mirrors M1 and M2 are aligned on the common

optical axis. The radii of curvature of both mirrors are optimized by local optimization in advance to correct the spherical aberration. Then, both mirrors have the radii $R_1 = -197.53\text{mm}$, $R_2 = -66.43\text{mm}$. The negative values show that both surfaces are concave to the incoming beam according to sign convention. The F-number in the image space is 2.2.

If both mirrors only rotate around the x -axis to avoid obscuration, the system is expanded in the y - z cross section. In this plane-symmetric system, the tangential cross sections of the two mirrors' local ray-incidence and local ray-caustic coordinate systems are coinciding with each other. The chief ray is oblique to both mirrors and results in different focal points for the tangential and sagittal aperture rays, which can be solved by the Coddington equations (2.9) and (2.10). The mirrors are in the air, which means

$$n = -n' = 1, I = -I' \quad (3.9)$$

Inserting Eq. (3.9) to the Coddington equations, one gets

$$\begin{cases} \frac{1}{s'} + \frac{1}{s} = \frac{2\cos I}{R_1} \\ \frac{1}{t'} + \frac{1}{t} = \frac{2}{R_1 \cos I} \end{cases} \quad (3.10)$$

Therefore, the focal lengths in the tangential and sagittal cross sections for the first mirror are

$$\begin{cases} f_s = \frac{R_1}{2} \cdot \frac{1}{\cos I} \\ f_t = \frac{R_1}{2} \cdot \cos I \end{cases} \quad (3.11)$$

where $0 \leq I \leq 90^\circ$, $R_1 < 0$. The expressions of the focal lengths for the second mirror can be achieved easily by substituting the subscript from 1 to 2. It can be easily derived from Eq. (3.11) that, for both mirrors, f_s must be smaller than f_t , which means that the contribution to astigmatism from both mirrors have the same sign. In other words, astigmatism cannot be compensated to zero no matter how the two concave mirrors tilt around the x -axis.

In a 3D Yolo telescope, M1 rotates around the x -axis so that it still has $f_s < f_t$ as before. However, the second mirror rotates around the y_c -axis of the local ray-caustic coordinate system about I degree as shown in Fig. 3-12. Because the incident chief ray (along the z_c -axis) and the surface normal of M2 \hat{e} are both perpendicular to the y_c -axis, the incidence plane is located in the x_c - z_c plane. The azimuthal angle θ between the local ray-incidence and local ray-caustic coordinate systems is 90° . By substituting $\eta = 0$ and $c_{\max} = c_{\min} = R_2$ for spherical surface into the generalized Coddington equations (2.12) to (2.14), the imaging equation for the second surface is

$$\begin{cases} \frac{1}{s'} + \frac{1}{s} = \frac{2}{R_2 \cos I} \\ \frac{1}{t'} + \frac{1}{t} = \frac{2 \cos I}{R_2} \end{cases} \quad (3.12)$$

which is exactly the opposite of Eq. (3.10). Therefore, one gets $f_s > f_t$ for the second mirror, which implies that the contribution of M2 to astigmatism has the opposite sign to M1. Then, by choosing proper tilting angles of both mirrors to adjust the incidence angles on the mirrors, the focal powers on both cross sections in the whole system can be the same, which is the condition that astigmatism vanishes.

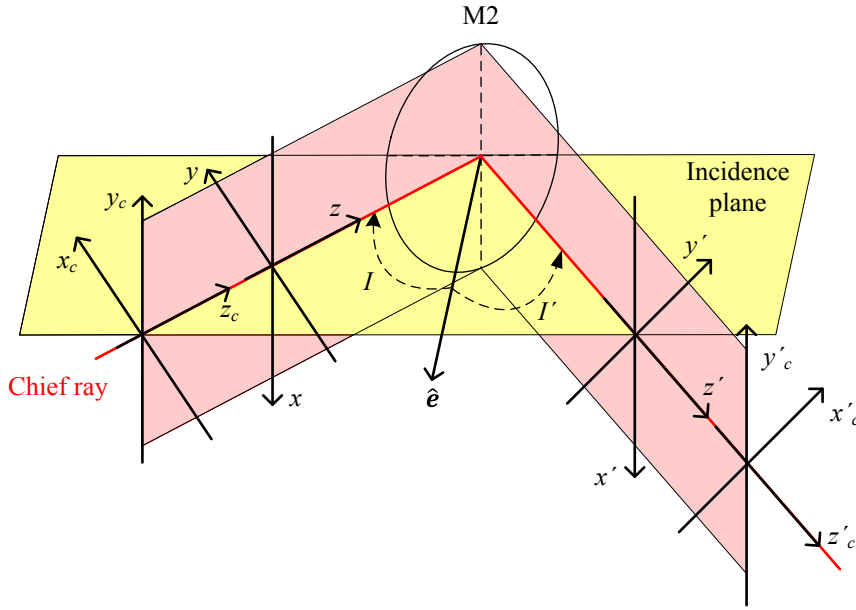


Fig. 3-12 M2 is rotated around y_c -axis. The azimuthal angles between the local ray-caustic coordinate system (x_c, y_c, z_c) and local ray-incidence coordinate system (x, y, z) in both object and image space are 90° .

After demonstrating the ability of the two-mirror Yolo telescope in correcting astigmatism, the next step is to eliminate both obscuration and astigmatism together by automatic optimization. In this example, the error function of the third order astigmatism is defined as

$$Erf_{asti} = \sqrt{Z_5^2 + Z_6^2} \quad (3.13)$$

where Z_5 and Z_6 are Zernike Fringe coefficients that represent the third order astigmatism. Then the total error function is rewritten as

$$Erf = Erf_{3D} + Erf_{asti} \quad (3.14)$$

which contains both obscuration and astigmatism parts. The tilt of M1 around the x -axis and the tilt of M2 around the y -axis are set as variables and they both have the range from -30° to 30° . Since there are two variables and one goal, Erf as a function of both variables are plotted in Fig. 3-13 by setting the clearance to be 0 and the bias to be 100mm^2 . The significant bias is to separate the obscured and unobscured optical structure with a

considerable gap. The part in the plot with a value larger than 100 is shown in yellow to represent that the system has obscuration when the mirrors move inside this area. The four corners with continuous change of the color bar show the unobscured areas with different astigmatism. Every corner has a dark blue valley where the primary astigmatism is minimized. It can also be noticed that the Erf plot is double-plane symmetric. There are two reasons: First, the geometry of the system is the same for the two opposite tilting angles; second, the $\cos I$ term is the same for the two opposite incidence angles in the Coddington equations.

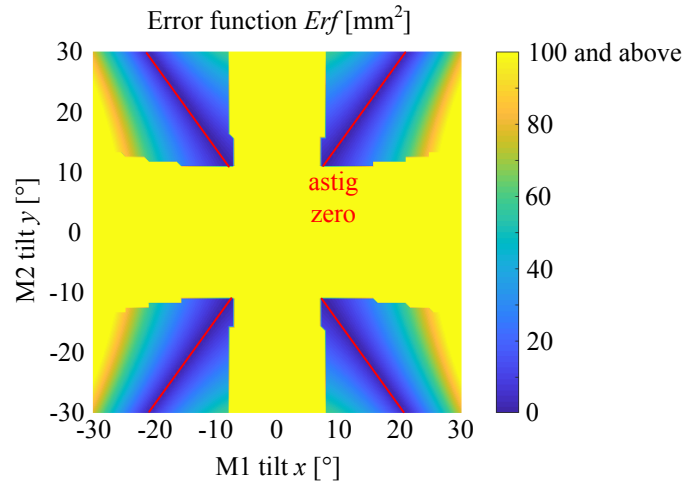


Fig. 3-13 Erf as a function of M1 tilt around x-axis and M2 tilt around y-axis. The value of Erf between 0 and 100 are mapped with color bar, all values larger than 100 are shown in the same color to represent obscured area. The red lines mark the valleys of zero astigmatism.

Again, the global optimization is utilized to search for the minimum. According to the symmetry property of the error function Erf , the range of the tilting angles of the two mirror are reduced to 0° to 30° . The bias is reduced to 5mm^2 for better continuity of the error function. The clearance is 0mm. The maximal iteration number is 200. After 200 iterations, as shown in Fig. 3-14, the global minimal error function is found (marked with the red circle) with the tilting angles (7.67° for M1, 11.11° for M2). After reaching this global minimum, the optimization tries to jump out the valley to search for another possibility with lower astigmatism. Nevertheless, it is trapped inside the unobscured part, which shows the benefits of bias. It should be mentioned that the error function does not go to zero in the end. This can be explained from the difference between Taylor expansion and Zernike Fringe polynomials. In Coddington equations, the optical surface is described by the 2nd order Taylor expansion. However, for Zernike Fringe polynomials, the quadratic terms are not only distributed in Z5 and Z6, but also appear in higher order terms like Z11, Z12 and so on. When the astigmatism is corrected, it means that the collection of all the Zernike Fringe terms that containing quadratic terms go to zero, but Erf_{asti} , which contains only Z5 and Z6, cannot reach zero alone. In addition, although the range of the variables is broad, the

3 General Solution for Obscuration Constraints

optimization result indicates that the optimization attempts to find the minimal rotating angles, which eliminate obscuration and astigmatism at the same time. The system layout after optimization is plotted in Fig. 3-15 (a) with no obscuration. The corresponding spot diagram is shown in Fig. 3-15 (b). It is obvious that the astigmatism is well corrected, and the residual aberration is dominated by coma.

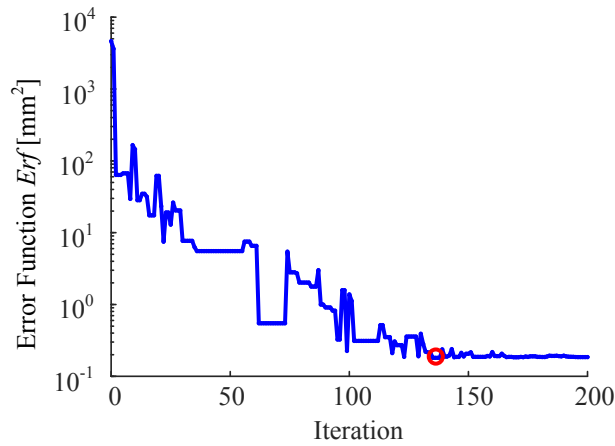


Fig. 3-14 Change of the error function *Erf* during optimization for 200 iterations. The best result region is marked with red circle.

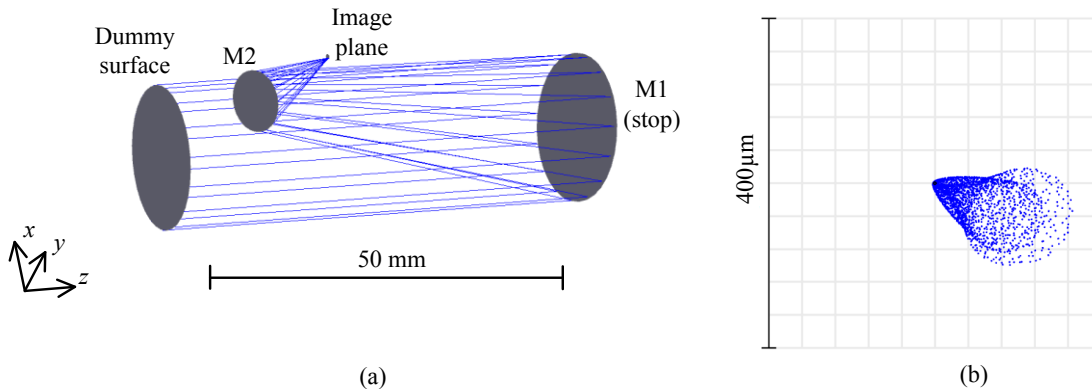


Fig. 3-15(a) Layout of unobscured two-mirror Yolo telescope after optimization. (b) The corresponding spot diagram.

Until now, the initial Yolo telescope, which is ready for further image quality optimization for a larger FOV, is found without human involvement. Then the FOV is set as $1^\circ \times 1^\circ$. Both mirrors are substituted by freeform surface – Zernike Fringe sag in OpticStudio. The conic parameters of both mirrors, the Zernike coefficients Z_5 to Z_9 of M1 and Z_5 to Z_{16} of M2 are set as variables. The two radii of curvature are fixed, otherwise, they are changing the system structure and may cause obscuration. The optimization criterion is the resolution of the whole FOV. The RMS spot radius map of the whole FOV before and after local optimization are drawn in Fig. 3-16. The averaged RMS spot radius is improved by 72.21 times, which shows the excellent correction capability of the freeform surface. The global tangential and sagittal view of the final Yolo telescope can be seen in Fig. 3-17.

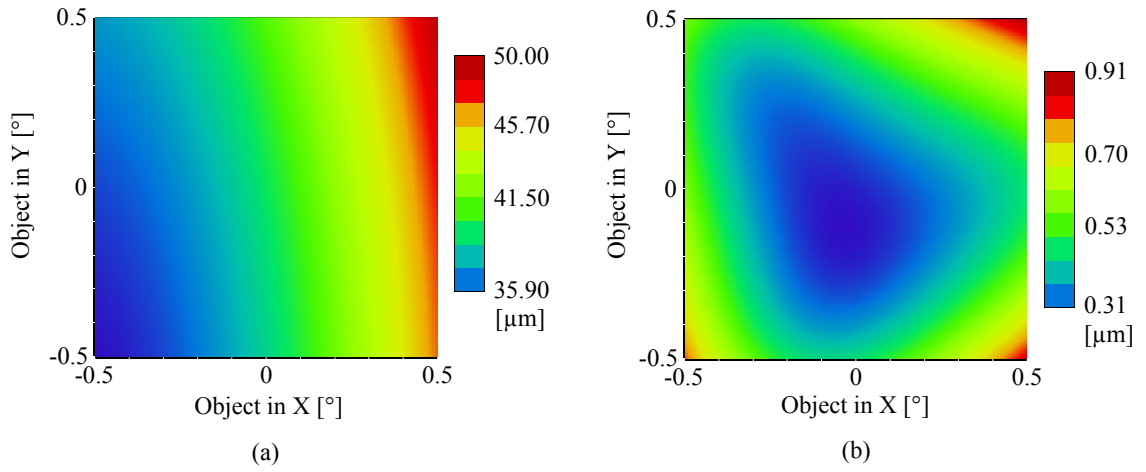


Fig. 3-16 The RMS spot radius map of the whole FOV before (a) and after (b) local optimization.

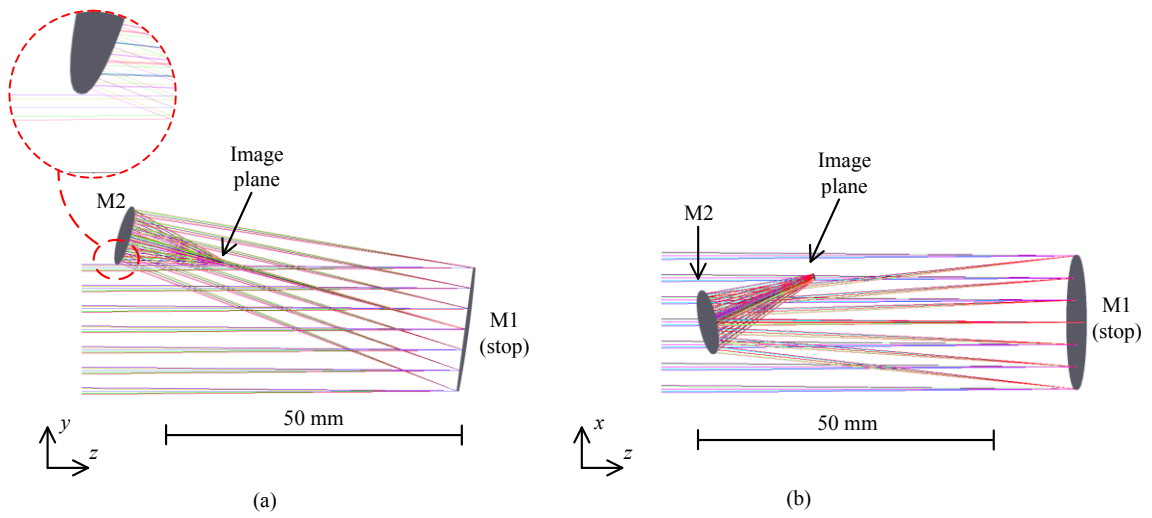


Fig. 3-17 The global tangential (a) and sagittal (b) view of the final Yolo telescope.

4 Quality Assessment of Chromatic Aberrations in Refractive Freeform Systems

4.1 Reference plane and reference axis

As mentioned in section 2.2, the design of non-rotationally symmetric refractive optical systems becomes popular in recent years. The propagation of the on-axis field in these systems has two main differences from the axisymmetric case. First, in general, the surfaces of non-rotationally symmetric optical systems can have different powers in the two perpendicular cross sections. It means that the paraxial ray bundle of the on-axis fields suffers from astigmatism and has a similar behavior as any other off-axis field. The generalized Coddington equations, which consider the propagation of paraxial astigmatic ray bundle around the chief ray through a single surface, should be applied to determine the tangential and sagittal image distances t' and s' . As a result, there is no unique Gaussian image plane in the intermediate image space. Also, from Eq. (2.12) to (2.14), it can be seen that s' and t' are wavelength-dependent, which means that the chromatic aberrations are coupled with monochromatic aberrations in the longitudinal direction. Secondly, if the refractive surfaces are allowed to rotate freely in free space, the OAR is no longer perpendicular to at least one of the optical surfaces. From the law of refraction, the deflection angle of the OAR is also wavelength-dependent. The OAR of varying wavelengths are separated after the first non-normal incidence surface. Consequently, there is no unique OAR. Based on the above two reasons, the chromatic aberrations in these optical systems cannot be treated by the Seidel aberrations, which does not take the astigmatic paraxial ray bundle into account, nor by Nodal aberration theory, which only considers monochromatic aberrations in off-axis optical systems. In order to evaluate the chromatic aberrations in general refractive optical systems, it is important to define the reference plane to calculate the chromatic aberration in the transverse direction and reference axis to calculate the chromatic aberration in the longitudinal direction.

No perfect image is obtained for a single field and a single wavelength due to the non-vanishing astigmatism. In Fig. 4-1, both image-sided tangential and sagittal paraxial coma ray paths are plotted on the same plane. The image location is defined at the circle of least confusion where the ray bundles in both cross sections have the same width h_c . According to the relation

$$h_c = (l' - t')NA'_t = (l' - s')NA'_s \quad (4.1)$$

The image distance of the circle of least confusion is obtained by

$$l' = \frac{s'NA'_s + t'NA'_t}{NA'_s + NA'_t} \quad (4.2)$$

which is depending on the individual values of s' , t' , and the image-sided numerical apertures NA'_s and NA'_t in the sagittal and tangential planes. By this approach, the widths of the ray bundle in the two cross sections are forced to be equal at the circle of least confusion; the ray fans in other transverse directions are not considered. In the case of macroscopic astigmatism, NA'_s and NA'_t are significantly different from each other, the real circle of least confusion becomes an “ellipse of smallest area”. Nevertheless, there is always a circle somewhere between the focal lines in our paraxial approach.

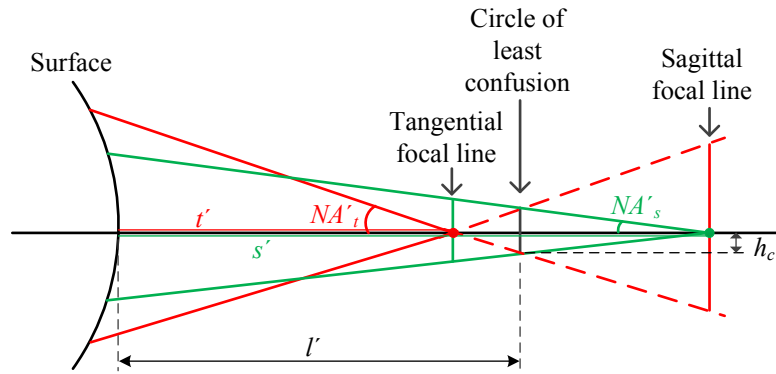


Fig. 4-1 Astigmatic imaging of a single field and a single wavelength.

If the tangential and sagittal focal lines, the circles of least confusion are calculated by the generalized Coddington equations and Eq. (4.2) for the whole FOV, then the tangential, sagittal and medial image shells in the image space for a single wavelength can be seen in Fig. 4-2. The paraxial image plane is defined at the circle of least confusion of the OAR and is tangent to the medial image shell in the paraxial region. The distance between the tangential and sagittal image shells is the astigmatism, which is now already finite on the OAR. The distance between the medial image shell and the paraxial image plane is the field curvature.

The central wavelength is usually selected as the reference wavelength for an optical system working in a broad-spectrum range. Accordingly, in the non-rotationally symmetric case, the paraxial image plane of the central wavelength is defined as the reference plane. Both OAR and CR of the central wavelength are defined as the reference axes for each object field individually because the change of astigmatism in the whole FOV is not symmetric around the on-axis field.

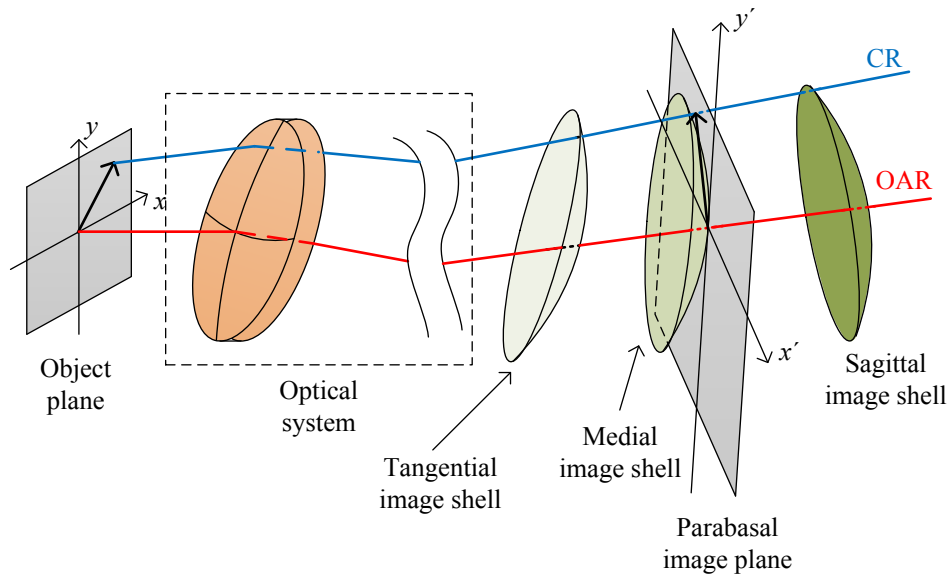


Fig. 4-2 Astigmatic imaging of the whole FOV for a single wavelength.

4.2 Definition of the generalized chromatic aberrations

In this section, the generalized chromatic aberrations are defined with regard to the reference plane and axis. Fig. 4-3 shows the optical paths of the chromatic OAR and CR in a refractive optical system of arbitrary symmetry. The OARs of the two outer wavelengths (λ_1 and λ_2) are separated after refraction at the first surface and thus induce decentered intersection points on the following surfaces as well as decentered image points, which are the effects that are different from the traditional situation.

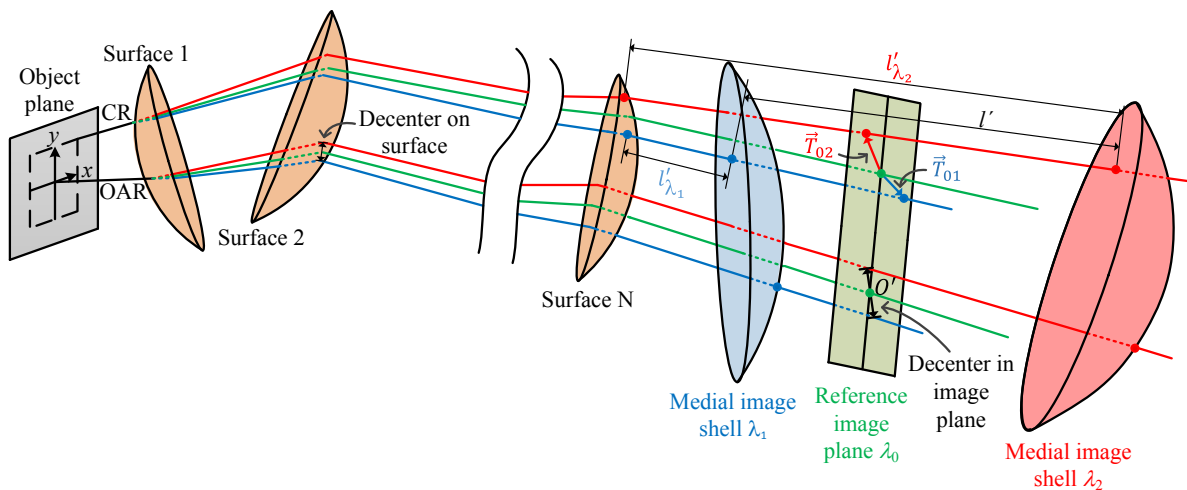


Fig. 4-3 The generalized chromatic aberrations in a non-rotationally symmetric refractive optical system.

For each field point, the image distance is measured from the ray intersection point on the surface to the circle of least confusion, for example l'_{λ_1} and l'_{λ_2} for the two outer wavelengths. The location of the circle of least confusion depends on the astigmatism of the corresponding field. Therefore, the chromatic aberration in the longitudinal direction is influenced by the

monochromatic aberrations. To distinguish from the longitudinal chromatic aberration in the traditional case, the generalized longitudinal chromatic aberration is called “chromatic image variation”. The aberrations along the longitudinal direction for different types of optical systems are compared in Table 4-1. The coupling between chromatic and monochromatic aberrations can also be seen for the off-axis fields in the rotationally symmetric optical system when the CRs are oblique to the optical surfaces. The chromatic image variation is defined as the distance between the medial image shells of the two out wavelengths along the chief ray path of the central wavelength λ_0 . The medial image shells are calculated by OAR or CR and parbasal coma rays. Therefore, the chromatic image variation is written as

$$l' = l'_{\lambda_2} - l'_{\lambda_1} \quad (4.3)$$

which is a scalar value that computed individually for each field. Although l'_{λ_1} and l'_{λ_2} are not exactly along the chief ray path of the central wavelength, the slight tilting angle can be neglected.

| | Monochromatic aberrations | Chromatic aberrations |
|--|----------------------------------|--|
| Axial field in rotationally symmetric system | None | Longitudinal chromatic aberration (analogous to defocus) |
| Off-axis field in rotationally symmetric system | Field curvature, astigmatism | Chromatic image variation |
| Fields in symmetry-free system | Field curvature, astigmatism | Chromatic image variation |

Table 4-1 Aberrations in the longitudinal direction for different types of optical systems.

The generalized transverse chromatic aberration is defined with regard to the parbasal image plane of the on-axis field with central wavelength λ_0 . One thing that should be noticed is that due to the possibility that the incidence plane rotates during ray propagation from surface to surface, the intersection points of the two outer wavelengths (λ_1 and λ_2) and the central wavelength λ_0 may not lie on one line in the image plane. The transverse chromatic aberration has to be calculated between λ_1 and λ_0 , λ_2 and λ_0 , respectively. As shown in Fig. 4-3, the transverse chromatic aberrations \vec{T}_{01} and \vec{T}_{02} are the vectorial difference of image intersection points between the two outer wavelengths and the central wavelength in the parbasal image plane of the central reference wavelength. Again, the difference between the generalized chromatic aberrations definitions in this thesis and the traditional definitions are compared in Table 4-2.

| | Traditional | | Generalized | |
|---------------------------|----------------------|----------------------------------|---|-----------------------------------|
| | Longitudinal | Transverse | Longitudinal | Transverse |
| Ray | Paraxial MR | Paraxial CR | CR, OAR & parabasal coma rays | CR & OAR |
| Intermediate image | Paraxial image point | Image plane perpendicular to OAR | Medial image shell constituted by circle of least confusion for individual field points | Parabasal image plane tilt to OAR |
| Decenter of OAR | No | No | Decenter on surface | Decenter in image plane |

Table 4-2 Comparison between traditional and generalized chromatic aberrations.

4.3 Ray-based calculation method

After defining the concepts of the generalized chromatic aberrations, it is essential to compute them quantitatively. The ray-based calculation method introduced in this section is a numerical-analytical hybrid method, in which the physical quantities of the OAR or CR are obtained by ray-tracing, the image shells and the angular displacements between different wavelengths are calculated and transferred to the image space analytically.

4.3.1 Chromatic image variation

In general, the parabasal tangential and sagittal focal line differences are considered separately between the two outer wavelengths behind each surface. The differences at each intermediate image space are transported to the final image space and accumulated to get the total contribution in both cross sections. Then, the difference of the circles of least confusion between the two wavelengths is computed to get the chromatic image variation.

Fig. 4-4 shows the propagation of the central ray (OAR or CR) starting from the object point *O* through the whole system to the image space. The locations of the sagittal focal lines are marked. The tangential focal lines are neglected in the plot because the principle in the tangential cross section is the same, only the focal distances are different.

For an arbitrary field and the two outer wavelengths (λ_1 and λ_2), the central ray is traced to get the object distance, incident and refracted angle, the local coordinate systems at each surface, respectively. The object distances for the two wavelengths before surface 1 are both

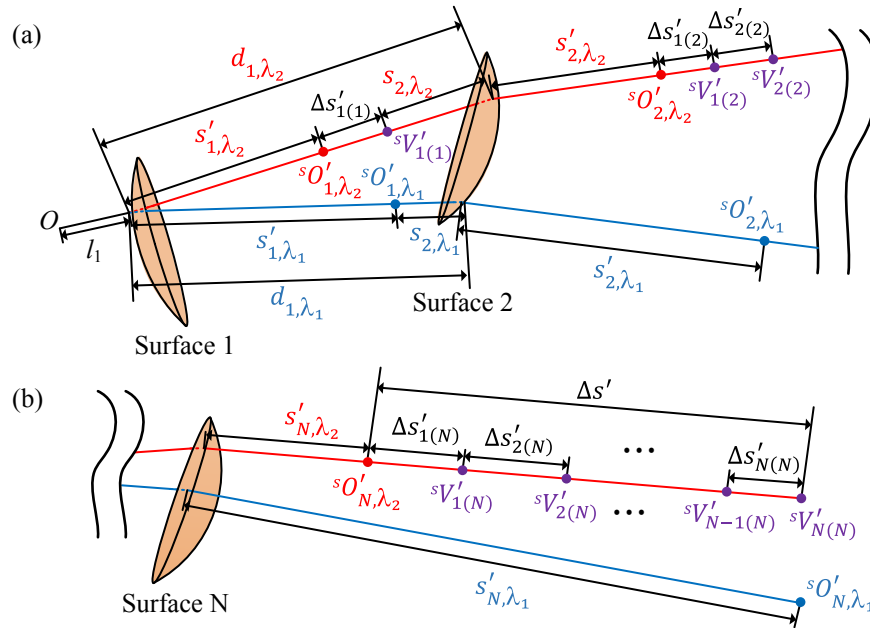


Fig. 4-4 Propagation of the chromatic intermediate image points of OAR or CR (a) from surface to surface and (b) in the final image space in the sagittal cross section. Blue and red indicate the physical quantities of λ_1 and λ_2 respectively. The purple points show the virtual image points and their transformation to the image space.

l_1 , the object-sided azimuthal angles are both 0. The image distances of the sagittal and tangential focal lines (s'_{1,λ_1} , t'_{1,λ_1} , s'_{1,λ_2} , t'_{1,λ_2}) and the image-sided azimuthal angles (θ'_{1,λ_1} , θ'_{1,λ_2}) are computed by the generalized Coddington equations (2.12) to (2.14). In the following, only the sagittal cross section is discussed, the formulae for the tangential cross section are the same and can be obtained by substituting “s” to “t”. The blue and red points sO'_{1,λ_1} and sO'_{1,λ_2} in Fig. 4-4 indicate the real sagittal foci. Another virtual sagittal image point $sV'_{1(1)}$ is then found on the refracted ray of λ_2 which has the same sagittal focal distance as λ_1 . The sagittal focal line difference generated by the first surface is

$$\Delta s'_{1(1)} = s'_{1,\lambda_2} - s'_{1,\lambda_1} = \overline{sO'_{1,\lambda_2} sV'_{1(1)}} \quad (4.4)$$

in which the subscript $a(b)$ means that the value of the physical quantity is generated by surface a and propagated to surface b . The superscript s on the left side denotes that the corresponding physical quantity belongs to the sagittal cross section.

After refraction at surface 1, the central rays are propagated to surface 2 to calculate the sagittal focal line difference generated by the second surface. The ray paths of the two wavelengths are separated now and intersect the following surfaces at different points, which cause an induced effect. In order to calculate intrinsic chromatic image variation of surface 2, the object points of both wavelengths should be sO'_{1,λ_1} and have the same object path as λ_1 . In order to take the induced effect under consideration, the object points of both wavelengths should be sO'_{1,λ_1} and $sV'_{1(1)}$, respectively, which have the same sagittal image

distance to surface 1 but different distances between the two surfaces (d_{1,λ_1} , d_{1,λ_2}) for the two rays. Therefore, the object distances for the two wavelengths are

$$s_{i,\lambda_1} = s'_{i-1,\lambda_1} - d_{i-1,\lambda_1}, s_{i,\lambda_2} = s'_{i-1,\lambda_2} - d_{i-1,\lambda_2} \quad (4.5)$$

where $i = 2, 3, \dots, N$ means that the relation is valid for the surfaces starting from the surface 2. The azimuthal rotation of the focal lines from the image space of surface $i-1$ to the object space of surface i are computed by

$$\theta_{i,\lambda_1} = \theta'_{i-1,\lambda_1} - \xi_{i,\lambda_1}, \theta_{i,\lambda_2} = \theta'_{i-1,\lambda_2} - \xi_{i,\lambda_2} \quad (4.6)$$

where ξ_i is the azimuthal angle between the incidence planes of surface $i-1$ and i . Then, the object point SO'_{i-1,λ_1} and $SV'_{i-1(i-1)}$ are imaged to SO'_{i,λ_1} and $SV'_{i-1(i)}$ by solving the generalized Coddington equations. Another virtual image point $SV'_{i(i)}$ is constructed by the same principle of $SV'_{1(1)}$. Then the sagittal focal line difference generated by surface i ($i \neq 1$) is

$$\Delta s'_{i(i)} = \overline{SV'_{i-1(i)} SV'_{i(i)}} \quad (4.7)$$

which contains both intrinsic contribution from the current surface and induced contribution from the former surfaces. In this way, the sagittal focal line differences generated by each surface are obtained. In the next step, the sagittal focal line differences are transferred to the image space. The simplest way is to differentiate the generalized Coddington equations to find the connection between the focal line difference in the object and image space. Then we can magnify all $\Delta s'_{i(i)}$ and $\Delta t'_{i(i)}$ to $\Delta s'_{i(N)}$ and $\Delta t'_{i(N)}$. However this method is only valid when $\Delta s'$ and $\Delta t'$ are both small comparing to s' and t' due to the natural discrepancy between derivative and finite difference. For an accurate calculation, all the virtual and real intermediate sagittal foci are transferred into the image space by solving the generalized Coddington equations. Accordingly, the sagittal focal line differences coming from all the surfaces in the image space are

$$\Delta s'_{i(N)} = \begin{cases} \overline{O_{N,\lambda_2}^s V_{1(N)}^s}, & \text{if } i = 1 \\ \overline{V_{i-1(N)}^s V_{i(N)}^s}, & \text{if } i = 2, 3, \dots, N \end{cases} \quad (4.8)$$

The total sagittal and tangential focal line differences are the summation of the surface contributions as

$$\Delta s' = \sum_{i=1}^N \Delta s'_{i(N)}, \Delta t' = \sum_{i=1}^N \Delta t'_{i(N)} \quad (4.9)$$

In the end, by differentiating Eq. (4.2), the relation among the total chromatic image variation and total sagittal and tangential focal line differences can be achieved as

$$\Delta l' = \sum_{i=1}^N \Delta l'_i = \sum_{i=1}^N \frac{NA'_s \cdot \Delta S'_{i(N)} + NA'_t \cdot \Delta t'_{i(N)}}{NA'_s + NA'_t} \quad (4.10)$$

which can also be written in the form of a surface summation.

4.3.2 Transverse chromatic aberration

After refraction at a single surface, the angular deflection of the chromatic refracted chief rays result in the transverse chromatic aberration in the image plane. Because of the rotation of incidence plane in the non-rotationally refractive optical systems, the transverse chromatic aberration is computed between the central wavelength λ_0 and the two outer wavelengths λ_1 and λ_2 , respectively. Fig. 4-5 shows the transformation of the transverse chromatic aberration between λ_0 and λ_2 generated at each surface to the image plane for an arbitrary field.

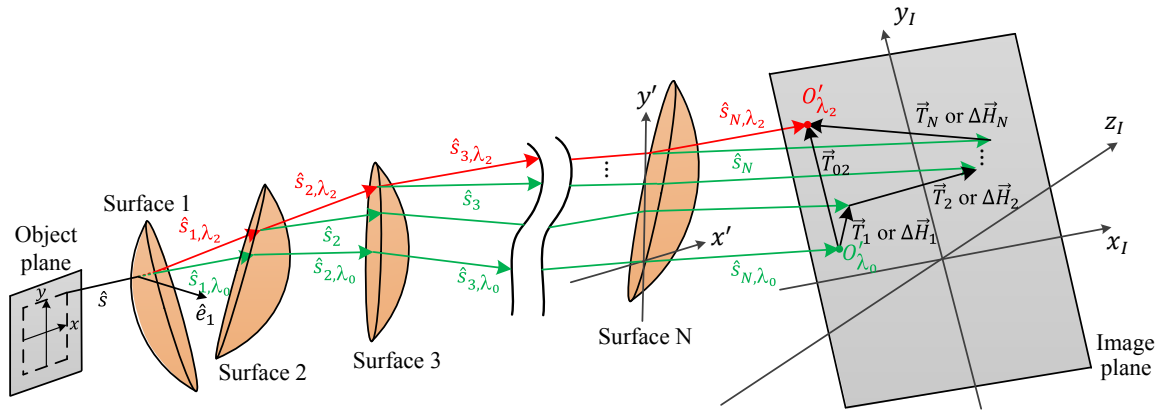


Fig. 4-5 Propagation of ray vectors which are generated at each surface to the final image plane to get the transverse chromatic aberration. $\Delta\vec{H}$ and \vec{T} are in the local ray-incidence coordinate system of the last surface and coordinate system in the image plane, respectively. The image plane is defined with respect to the central wavelength λ_0 .

Both wavelengths have the same direction cosine \hat{s} in the object space and are incident at the same point on surface 1. The surface normal at the incidence point is \hat{e}_1 . Using the law of refraction in the vectorial form

$$n' \hat{s}' = n \hat{s} + \hat{e} \left[\sqrt{n'^2 - n^2 + (n \hat{e} \cdot \hat{s})^2} - n \hat{e} \cdot \hat{s} \right] \quad (4.11)$$

the direction cosines \hat{s}_{1,λ_0} and \hat{s}_{1,λ_2} of both wavelengths behind the first surface are obtained. The angular difference between \hat{s}_{1,λ_0} and \hat{s}_{1,λ_2} is the contribution from the first surface.

Then, green and red chief rays intersect surface 2 and the following surfaces at different points. The angular dispersion at the green chief ray intersection point is the intrinsic contribution from the current surface. The angular dispersion at the red chief ray intersection point includes both intrinsic and induced contributions. In the following, the steps to compute the total contribution of a surface are introduced. The calculation principle of the intrinsic aberration is the same, only the incident ray is different. Starting from surface 2, the

dispersion at the outer wavelength intersection point is calculated by using \hat{s}_{i-1,λ_2} as the object-sided direction cosine and substituting n and n' of λ_0 and λ_2 respectively to Eq. (4.11). \hat{s}_i shows the virtual ray path of λ_0 refracted at the red chief ray intersection point on the surface. Then the angular difference between \hat{s}_{i,λ_2} and \hat{s}_i is the overall contribution from surface i .

Next, the surface contributions are transported to the final image space by linear transformation. The ray path described by the direction cosine $\hat{s}_i = (\alpha_i, \beta_i, \gamma_i)$ behind the surface i is governed by the ray vector $\Delta\vec{r}_i$ as [2-1]

$$\Delta\vec{r}_i = \begin{bmatrix} x_i \\ y_i \\ U_i \\ V_i \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \\ \frac{\alpha_i}{\gamma_i} n'_{i,\lambda_0} \\ \frac{\beta_i}{\gamma_i} n'_{i,\lambda_0} \end{bmatrix} \quad (4.12)$$

where x_i and y_i are the ray intersection point components on surface i in the local ray-incidence coordinate system (x', y', z') , U_i and V_i are the numerical aperture of the ray in the $x'z'$ - and $y'z'$ - planes, n'_{i,λ_0} is the refractive index of λ_0 behind the surface. The ray vectors $\Delta\vec{r}_2, \Delta\vec{r}_3, \dots, \Delta\vec{r}_{N-1}$ are transferred to the final image space by the multiplication of the 4×4 transformation matrix \mathbf{M}_i at the following surfaces. The details of the transformation matrix are explained in Appendix A. Therefore, the ray vectors behind the last surface are

$$\Delta\vec{r}'_i = \begin{bmatrix} x'_i \\ y'_i \\ U'_i \\ V'_i \end{bmatrix} = \begin{bmatrix} x'_i \\ y'_i \\ \frac{\alpha'_i}{\gamma'_i} n'_{N,\lambda_0} \\ \frac{\beta'_i}{\gamma'_i} n'_{N,\lambda_0} \end{bmatrix} = \begin{cases} \mathbf{M}_N \cdot \mathbf{M}_{N-1} \cdots \mathbf{M}_{i+1} \cdot \Delta\vec{r}_i, & \text{if } i = 2, 3, \dots, N-1 \\ \Delta\vec{r}_i, & \text{if } i = N \end{cases} \quad (4.13)$$

which is formulated in the local ray-incidence coordinate system of the last surface.

The ray vectors give the information of the relative ray height and direction in the image space. Because the paraxial image plane is tilted to the reference chief ray, the ray intersection point on the image plane has to be calculated for each ray vector with their own propagation distance d'_i as

$$(x''_i, y''_i, z''_i) = (x'_i + d'_i \cdot \alpha'_i, y'_i + d'_i \cdot \beta'_i, z'_i + d'_i \cdot \gamma'_i) \quad (4.14)$$

Combining the real chief ray intersection points of both wavelengths $O'_{\lambda_0} = (x''_1, y''_1, z''_1)$ and $O'_{\lambda_2} = (x''_{N+1}, y''_{N+1}, z''_{N+1})$ in the image plane, the surface-by-surface transverse chromatic aberration in the local ray-incidence coordinate system of the last surface can be written as

$$\Delta\vec{H}_i = [x''_{i+1} - x''_i, y''_{i+1} - y''_i, z''_{i+1} - z''_i], i = 1, 2, \dots, N \quad (4.15)$$

The transverse chromatic aberrations are transformed to the coordinate system of the image plane (x_l, y_l, z_l) by

$$\vec{T}_i = \mathbf{G} \cdot \Delta\vec{H}_i^T \quad (4.16)$$

in which \mathbf{G} is the rotational matrix between the coordinate systems (x', y', z') and (x_l, y_l, z_l) . Now, the z component in \vec{T}_i is exactly zero. The total transverse chromatic aberration is

$$\vec{T}_{02} = \sum_{i=1}^N \vec{T}_i \quad (4.17)$$

4.4 Wavefront-based calculation method

In this section, the generalized chromatic aberrations are calculated from the viewpoint of wavefront expansion. As introduced in section 2.4, the wavefront aberration is deduced from the optical path difference between the real rays and the ideal rays. Considering the chromatic effect for a single field, the ideal ray can be defined as the chief ray of the reference wavelength; the real rays can be defined as the chief ray and paraxial comas rays of another wavelength. By this approach, the chromatic image variation and the transverse chromatic aberration can be viewed as the defocus between the circles of least confusion of the two wavelengths and angular tilt between the chief rays of the two wavelengths. The real and ideal wavefronts are constructed in section 4.4.1. The additivity of the wavefront is demonstrated in section 4.4.2. The wavefront is fitted to obtain the chromatic aberrations in the last section.

4.4.1 Real wavefront and reference sphere

The paraxial ray bundle starting from an object point O is propagated to surface N as shown in Fig. 4-6. The chromatic ray paths are separated during propagation. As a result, the chief rays of the reference wavelength and the other wavelength intersect the surface at the points \bar{P}_0 and \bar{P} , respectively. The circles of least confusion of the two wavelengths are calculated by the generalized Coddington equations with the locations at O'_0 and O' . The tangential ray bundle is drawn and focused at the point O'_t . The sagittal ray bundle is omitted, but the sagittal focal point O'_s is shown.

Typically, the reference sphere is constructed relative to the reference chief ray at the exit pupil. In order to simplify the calculation process, the reference sphere here is built with the vertex on the surface. The reason to do so is that each field is considered individually. There is no need to consider the common stop for the whole FOV. For each field, it can be assumed

that the stop is located at the intersection of the reference and real chief ray, which is close to the surface. Since the influence of varying radii of the reference sphere is of higher order [4-1], the small shift between the stop and the surface can be neglected. Otherwise, huge calculation is necessary and time-consuming, especially when the number of surfaces is large.

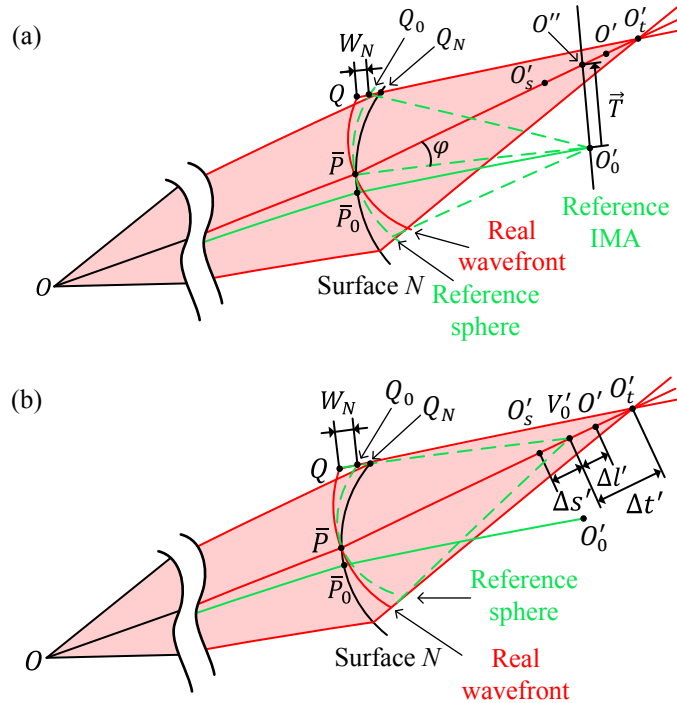


Fig. 4-6 Construction of reference sphere for (a) tilt and (b) defocus.

Two types of the generalized chromatic aberrations have to be discussed separately because the reference spheres in both cases are constructed differently due to the displacement of the chromatic chief rays on the surface. For the transverse chromatic aberration shown in Fig. 4-6 (a), the reference sphere is drawn as green dashed curve with vertex at the point \bar{P} and radius equal to $\bar{P}O'_0$. The real chief ray intersects the reference image plane at the point O'' . According to the definition, the total transverse chromatic aberration behind surface N is $|\vec{T}| = O'_0O''$, and the tilt angle is φ . Under this configuration, the defocus is the difference between $\bar{P}O'$ and $\bar{P}O'_0$. However, the chromatic image variation according to the definition should be the difference between $\bar{P}O'$ and $\bar{P}_0O'_0$. Since $\bar{P}O'_0$ and $\bar{P}_0O'_0$ are not exactly equal, another configuration as shown in Fig. 4-6 (b) is established to calculate the chromatic image variation. In this setup, a point V'_0 is found on the real chief ray to fulfill the requirement that the radius of the reference sphere $\bar{P}V'_0$ equals $\bar{P}_0O'_0$. Finally, $\Delta s'$ and $\Delta t'$ are the sagittal and tangential defocus, and $\Delta l'$ is exactly the chromatic image variation.

The wavefront aberration consists of the optical path differences between the real and ideal rays. For an arbitrary real ray, for example, the real tangential coma ray with the optical path $[OQ_NO'_t]$ is considered, its optical path from O to the real wavefront is $[OQ]$. Q_N is the ray

intersection point on the surface. The ideal ray path to the reference sphere is $[OQ_0]$. Q_0 is the intersection point between the refracted ray $Q_N O'_t$ and the reference sphere. The optical path difference for the real tangential coma ray is computed as

$$OPD = [OQ] - [OQ_0] = [OP] - [OQ_0] \quad (4.18)$$

which can be input into Eq. (2.15) to calculate the wavefront aberration W_N . [...] represents the optical path length. $[OQ]$ is substituted to $[OP]$ because all the points on the same wavefront have the same optical path length.

4.4.2 Additivity of phase

Using the method introduced in the former section, the chromatic wavefront aberration W_i behind every surface can be achieved as shown in Fig. 4-7, which is the accumulated wavefront deviation contributed from the first surface until surface i . The total chromatic wavefront aberration of the system is W_N . The chromatic wavefront aberration ΔW_i generated by a single surface i has the following relation

$$\Delta W_i = \begin{cases} W_1, & \text{if } i = 1 \\ W_i - W_{i-1}, & \text{if } i = 2, 3, \dots, N \end{cases} \quad (4.19)$$

Accordingly, the total chromatic wavefront aberration can be rewritten as

$$W_N = W_1 + (W_2 - W_1) + \dots + (W_N - W_{N-1}) = \sum_{i=1}^N \Delta W_i \quad (4.20)$$

It shows that the total chromatic wavefront aberration is the summation of single surface contribution. It proves mathematically that the generalized chromatic aberrations are additive.

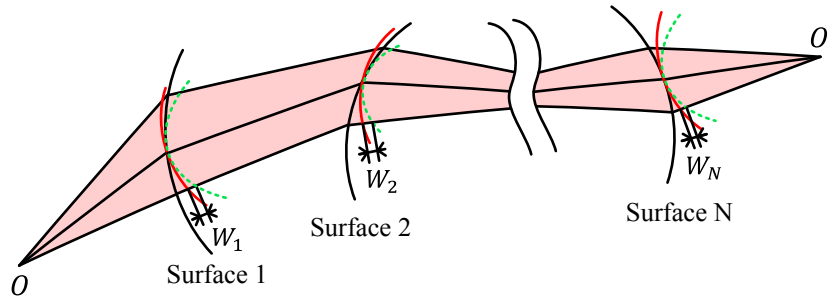


Fig. 4-7 Total chromatic wavefront aberration behind each surface.

4.4.3 Fitting of the chromatic aberrations

Next step is to expand the wavefront aberration into defocus and tilt and to derive the chromatic image variation in the longitudinal direction and transverse chromatic aberration in the lateral directions from the mathematical relation between the wave and ray aberrations.

The optical path of a ray can be obtained by finite raytracing. The circle of least confusion, tangential and sagittal focal lines are computed analytically. Since the aberrations in the

tangential and sagittal cross sections are different, and both defocus and tilt are of the 2nd order, the wavefront aberration behind surface i is expanded up to the second order in both x - and y -pupil coordinates as follow

$$W_i(x_p, y_p) = \varepsilon_1 x_p + \varepsilon_2 y_p + \varepsilon_3 x_p^2 + \varepsilon_4 y_p^2 + \varepsilon_5 x_p y_p \quad (4.21)$$

where ε_1 to ε_5 are the coefficients that need to be fitted, x_p and y_p are tilt terms, x_p^2 and y_p^2 are defocus terms, $x_p y_p$ shows the rotation of focal lines during propagation. There are in total five terms. Therefore, the information of five rays is demanded to get the unique fitting result. The five rays are chosen as the chief ray, upper and lower sagittal and tangential parabolal coma rays of the wavelength under consideration. If the pupil coordinates are the same as the local ray-caustic coordinate system at each surface, the normalized x_p and y_p for these five rays on each surface are always (0, 0), (1, 0), (-1, 0), (0, 1), (0, -1). The term $x_p y_p$ is therefore zero. If the pupil coordinates coincide the local ray-incidence coordinate system at each surface, the $x_p y_p$ indicates the azimuthal rotation of astigmatism. After fitting, the five terms in Eq. (4.21) are decoupled.

The relation between the wave aberration and transverse aberration for the tilt term is

$$\begin{cases} W_{tilt_x} = \varepsilon_1 x_p = -\frac{n' x_p}{R_{ref}} \Delta x' = -NA'_s \Delta x' \\ W_{tilt_y} = \varepsilon_2 y_p = -\frac{n' y_p}{R_{ref}} \Delta y' = -NA'_t \Delta y' \end{cases} \quad (4.22)$$

where R_{ref} is the radius of the reference sphere, $\Delta x'$ and $\Delta y'$ are the x - and y -components of transverse chromatic aberration on the local image plane with reference to the chief ray of individual field. The final transverse chromatic aberrations have to be projected onto the common image plane, which is the parabolal image plane of the OAR for the central wavelength.

The relation between the wave aberration and longitudinal aberration for the defocus term is

$$\begin{cases} W_{defocus_x} = \varepsilon_3 x_p^2 = -\frac{n'}{2} \left(\frac{x_p}{R_{ref}} \right)^2 \Delta s' = -\frac{1}{2n'} NA'_s{}^2 \Delta s' \\ W_{defocus_y} = \varepsilon_4 y_p^2 = -\frac{n'}{2} \left(\frac{y_p}{R_{ref}} \right)^2 \Delta t' = -\frac{1}{2n'} NA'_t{}^2 \Delta t' \end{cases} \quad (4.23)$$

After obtaining $\Delta s'$ and $\Delta t'$, the chromatic image variation $\Delta l'$ can be obtained by Eq. (4.10).

4.5 Influence of chromatic pupil aberration

The pupil discussed in this section is not the local stop on the individual chief ray to construct the reference sphere as mentioned in section 4.4.1, but the stop of the whole system. In an

optical system suffering from chromatic aberrations, the chromatic pupil aberrations also exist because the object and the pupil imaging conditions can change from the one to the other by the exchange of MR and CR. In the former discussion, the chief rays of different colors have the same ray path in the object space when they are starting from the same object point. This is true only if the stop is in the object space, such as scanning systems or some cell phone camera systems. In most optical systems, the stop is located somewhere inside the system. Due to the chromatic pupil aberration, the entrance and exit pupils (EnP and ExP) of different wavelengths are displaced along the axis as shown in Fig. 4-8, which is similar to the “pupil walking” effect in the monochromatic aberrations [4-1]. But the reason for the pupil shifting effect is different between chromatic and monochromatic aberrations. In the chromatic case, the pupil shifts because of the dispersion of material. However, in the monochromatic case, the pupil shifts due to the aberration of chief ray. As a consequence of the entrance pupil displacement, the ray paths of the chromatic chief rays from the same object point are different, and should be taken into consideration during calculation.

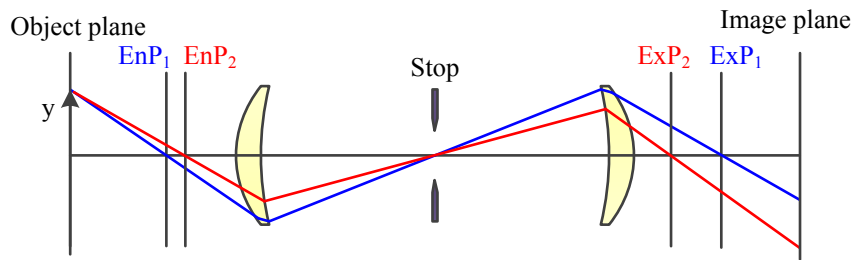


Fig. 4-8 Displacement of pupils between two wavelengths due to the chromatic pupil aberration.

Fig. 4-9 shows the separated chief ray paths in the object space for two wavelengths. Now the two wavelengths have their own object distances, incidence angles, direction cosines and surface normals, but the azimuthal angles are still both zero because the tangential plane of the ray bundle coincides with the incidence plane in the object space. When calculating the generalized Coddington equations and the vectorial refraction equation, the corresponding physical quantities must be substituted for the two wavelengths individually. In the wave-

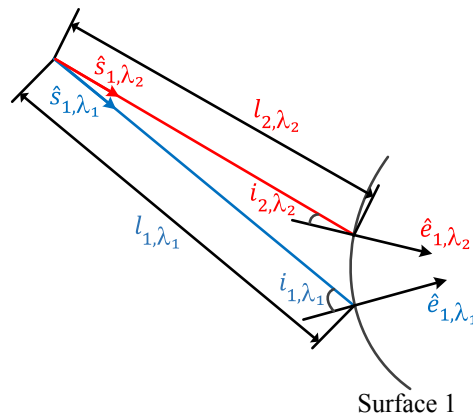


Fig. 4-9 Separated chief ray paths for two wavelengths and their own physical quantities.

front-based method, the modified ray paths are traced due to the chromatic pupil aberration, the chief ray displacement on the first surface is then automatically included.

4.6 Examples

In the remaining part of this chapter, four examples are given to verify the ray-based and wavefront-based calculation methods and the unification of the definition of the generalized chromatic aberrations for the optical systems with arbitrary symmetry. In order to present and evaluate the generalized chromatic aberrations in a meaningful way, the chromatic image variation is normalized by the diffraction limit criterion in the longitudinal direction, which is the Rayleigh unit (RL) or also called as the depth of focus of the central wavelength. Correspondingly, the Airy diameter (D_{Airy}) of the central wavelength, which shows the diffraction limit condition in the transverse direction, is used to normalize the transverse chromatic aberration.

4.6.1 Double Gauss system

Double Gauss system is a group of photographic systems with two quasi-symmetric lens groups and a stop in the middle. The quasi-symmetric structure of the systems ensures the excellent suppression of the odd order aberrations, such as coma, distortion and lateral color. The layout of a double Gauss example system provided by OpticStudio [4-2] is shown in Fig. 4-10. Dark and bright orange indicate different flint glasses, and yellow shows the crown glass. Table 4-3 lists the basic data. The double Gauss system is rotationally symmetric. Thus, the second order longitudinal and transverse chromatic aberrations are well described by the Seidel aberrations. Since the generalized chromatic aberrations are of full orders and generated for systems with arbitrary symmetry, they are also compatible with the rotationally symmetric system and should give the same results as the Seidel aberrations in the paraxial region where the influence of higher order chromatic aberrations can be neglected. Therefore, the object field to calculate transverse chromatic aberration is set as 0.5° .

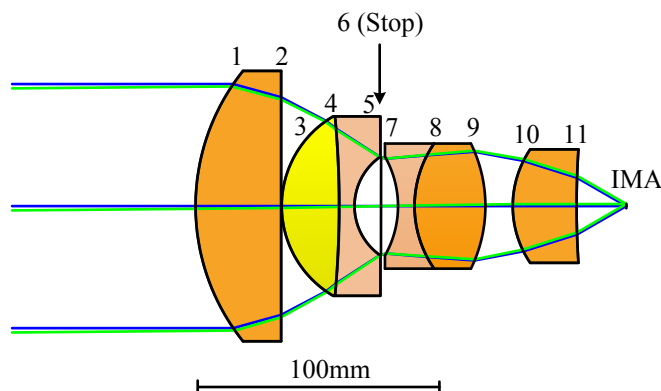


Fig. 4-10 Layout of the double Gauss system.

| Parameters | Values |
|--------------|-------------|
| Focal length | 101.6mm |
| F-number | 1 |
| Wavelengths | F'eC'-lines |

Table 4-3 Basic data of the double Gauss system.

The surface-by-surface longitudinal and transverse chromatic aberrations of the double Gauss system are shown in Fig. 4-11. Ray-based and wavefront-based results are calculated from the methods introduced in sections 4.3 and 4.4. The results of Seidel aberrations are read from OpticStudio. From the bar graph, the good consistency among the three methods can be clearly seen despite some minor differences. The results prove the convergence of the definitions given in this thesis for the special case of circular symmetry. The reason for the slight difference is the nonlinearity of the material dispersion. The ray- and wavefront-based methods use the exact values of the refractive indices to calculate the chromatic aberrations. However, referring back to section 2.4.1, Seidel derived the chromatic aberrations by differentiating the refractive index as a function of wavelength, in which the non-linear dispersion curve of the material is approximated to be linear. In addition, the bar graph proves that the transverse chromatic aberrations can be compensated by the quasi-symmetric

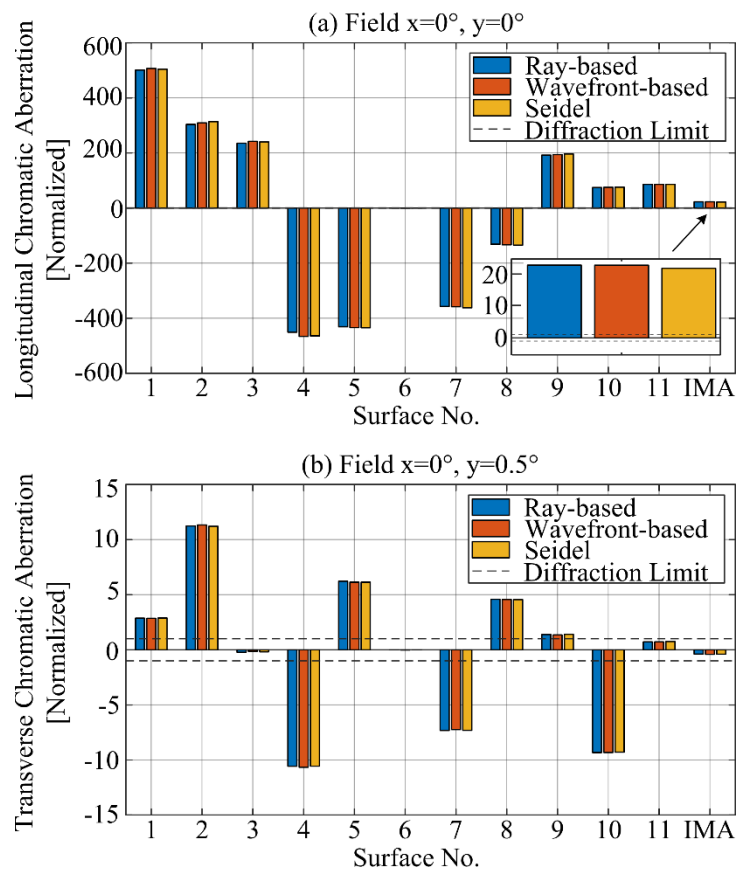


Fig. 4-11 Surface-by-surface comparison of (a) the longitudinal chromatic aberration and (b) the transverse chromatic aberration generated by the three methods.

structure of the double Gauss system because the surface contributions are nearly symmetric around the stop (surface 6) with the opposite sign and the total transverse chromatic aberration is diffraction limited. However, the residual longitudinal chromatic aberration is still dramatic.

4.6.2 Stereomicroscope

The second example is a simple setup of a stereomicroscope system, which can be seen from Fig. 4-12. It contains three lens groups - an objective lens for imaging, a Galilean telescope for change of magnification, and a tube lens for refocusing. The objective lens with a rear freeform surface is decentered to allow for two channels for binocular vision. The basic data is listed in Table 4-4.

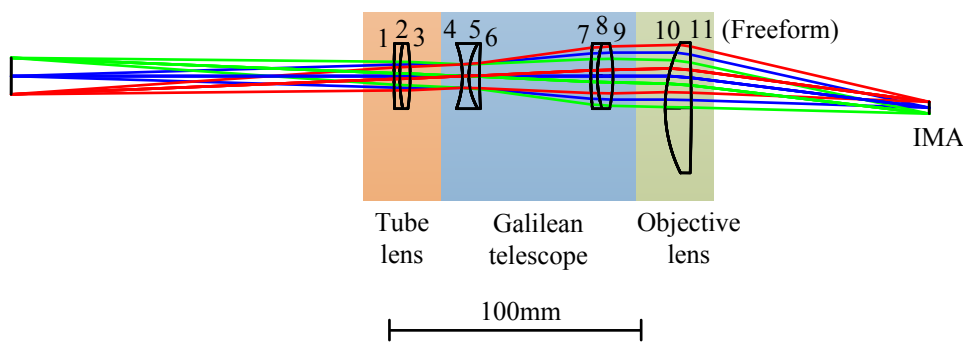


Fig. 4-12 Layout of the stereomicroscope for one eye channel.

| Parameters | Values |
|-----------------|-------------|
| Focal length | 60mm |
| FOV (x×y) | 14mm×14mm |
| Object-sided NA | 0.03 |
| Wavelengths | F'eC'-lines |

Table 4-4 Basic data of the stereomicroscope.

First, the surface contributions to the chromatic image variations in this system are calculated by both ray- and wavefront-based methods and plotted in Fig. 4-13 for the maximum field corner $x, y=7\text{mm}$. It can be seen that the results obtained by both methods are nearly identical. The slight differences are due to the higher order effects caused by changing the radii of the reference spheres behind each surface in the wavefront-based method. The line chart shows the accumulated chromatic image variation behind each surface. Apparently, the chromatic image variations are well corrected in the first two lens groups because of the utilization of the achromats. The major chromatic image variation is caused by the single objective lens.

The surface contributions to the transverse chromatic aberrations between the eF'-lines are also calculated by the two methods for the same field corner and compared in Fig. 4-14. Because the transverse chromatic aberrations are vectors, the x - and y -components are

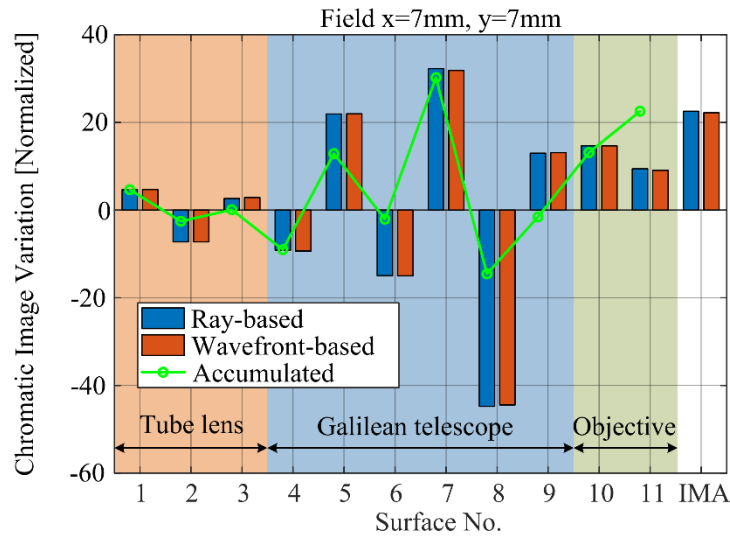


Fig. 4-13 Surface-by-surface chromatic image variations calculated by ray- and wavefront-based methods.

plotted, respectively. The quantitative values for both components can be found in Appendix B. In the rotationally symmetric optical system, the transverse chromatic aberration of the field point with equal x - and y -heights has the same amplitude in x - and y -components. It is also true in this stereomicroscope for the transverse chromatic aberrations at the first 9 spherical surfaces, which are aligned on the same optical axis. Therefore, only the x -components are plotted for these surfaces. The minor difference between the x - and y -directions is due to the asymmetric lateral magnification coming from the decentered objective lens. From the accumulated line chart, the transverse chromatic aberrations are also well corrected in the first two lens groups. The decentered objective lens results in unequal values and opposite signs between the two components.

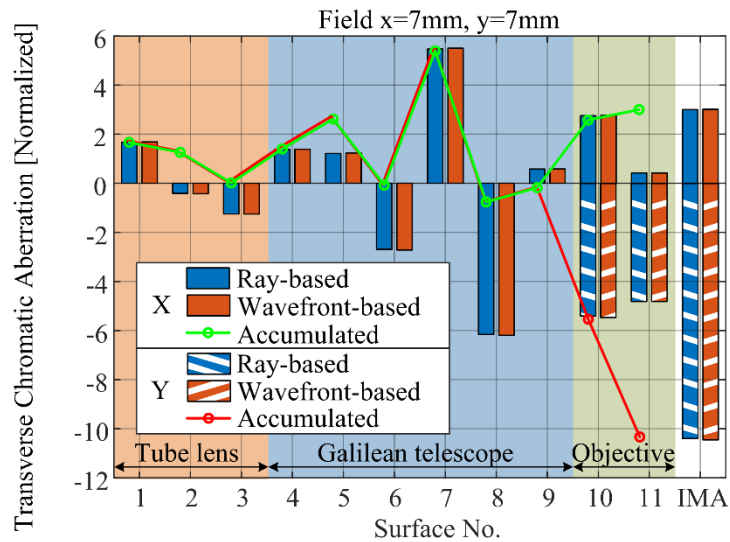


Fig. 4-14 Surface-by-surface transverse chromatic aberrations between the eF' -lines calculated by ray- and wavefront-based methods.

Since the distribution of the generalized chromatic aberrations is no longer symmetric around the on-axis field for the whole FOV, the full-field display is helpful to analyze the

chromatic performance among different field points. In Fig. 4-15, the surface-by-surface chromatic image variations of the OAR are chosen as the nominal values. The full-field displays of the chromatic image variations in surfaces 8, 11, and image plane are plotted as the difference to the nominal values. Surface 8 is inside the rotationally symmetric configuration, and the CR paths are rotationally symmetric around the OAR. However, due to the asymmetric axial magnification introduced by the off-axis objective lens and the freeform surface, the full-field contribution from surface 8 is no longer rotationally symmetric. The fields, which have the same contribution as the on-axis field, are located on the dashed curve. The fields above (or below) the curve have smaller (or larger) contributions than the on-axis field. The full-field distribution on surface 11 behaves different. Because of the positive and negative compensation among the surfaces, the full-field distribution in the image plane is nearly rotationally symmetric.

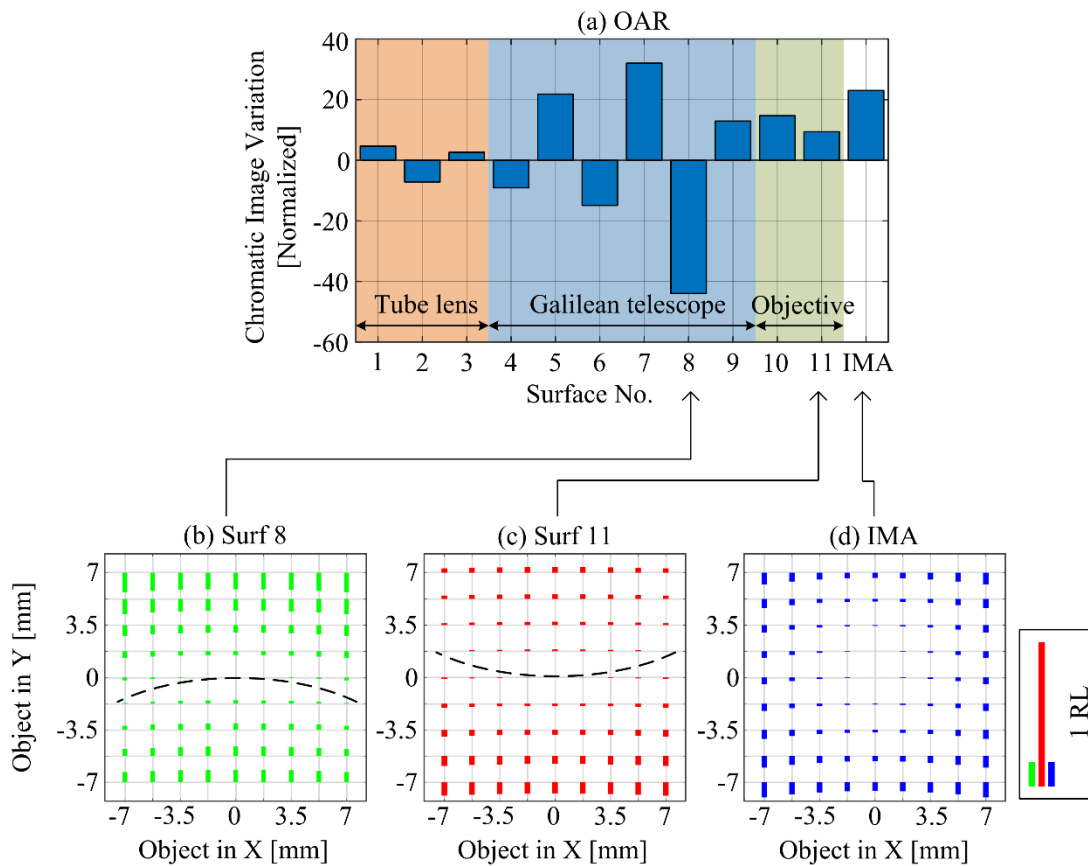


Fig. 4-15 (a) Surface-by-surface chromatic image variations of the on-axis field. The full-field display of surfaces (b) 8, (c) 11 and (d) image plane are shown as the difference between the on-axis field and other fields. The dashed curves mark the fields, which have the same contribution as the on-axis field.

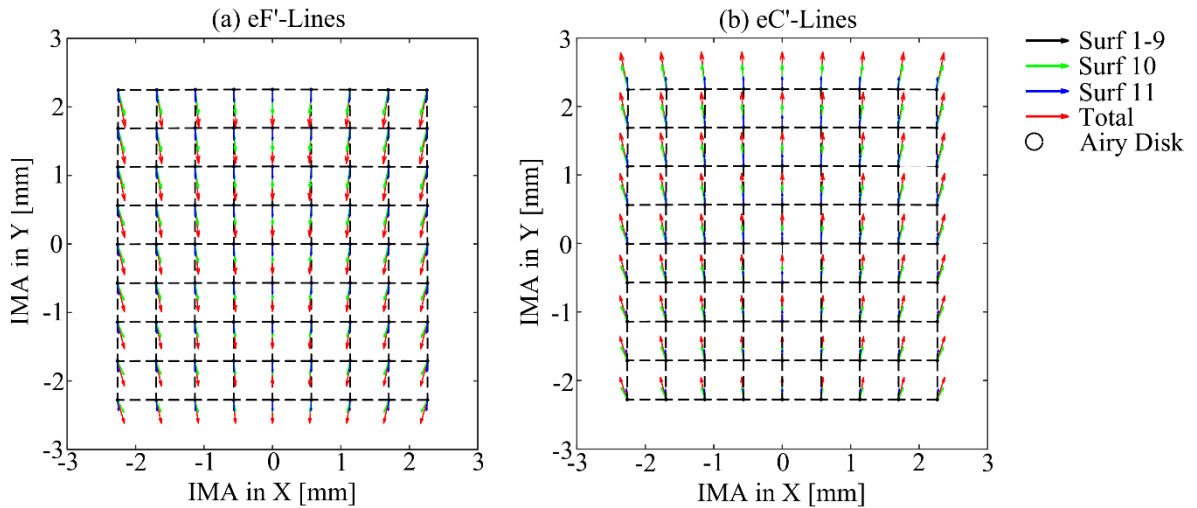


Fig. 4-16 Full-field display of the transverse chromatic aberration between the (a) eF²-lines and (b) eC²-lines.

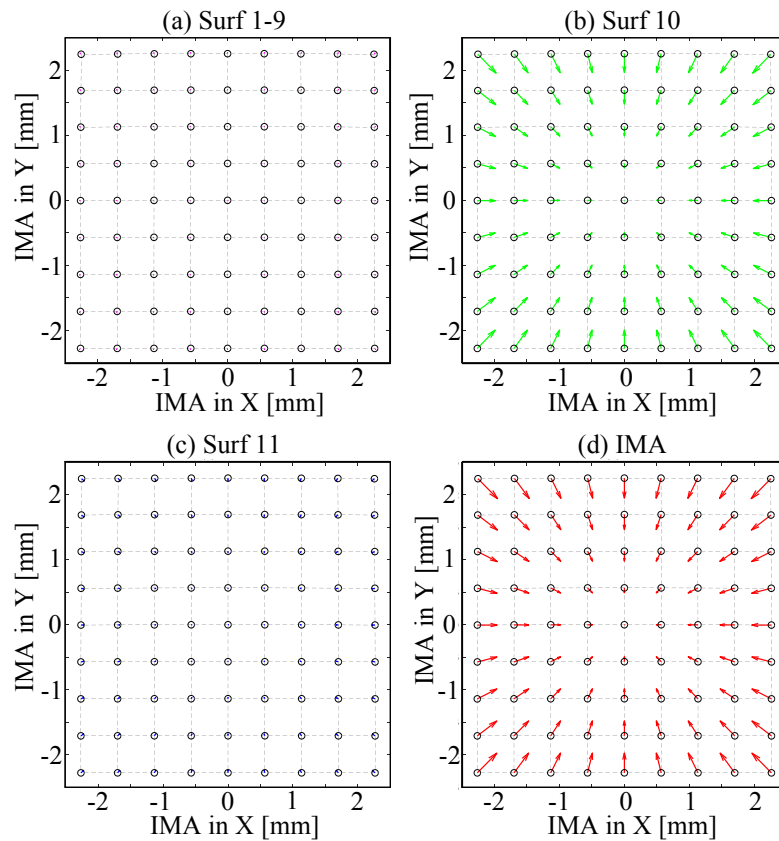


Fig. 4-17 Full-field display of the transverse chromatic aberration between the eF²-lines after removing the decentering effect for (a) the summation of surfaces from 1-9, (b) surface 10, (c) surface 11 and (d) the image plane.

The full-field display of the transverse chromatic aberrations is plotted for eF²-lines and eC²-lines separately in Fig. 4-16 because the two arrows indicating the contribution from the same surface in the two spectral ranges are not aligned in the non-rotationally symmetric systems, although it is not evident in this example. Also, for a single field point, the arrows of different surface contributions are not on the same line, which shows the asymmetric property of the system. The accumulated contribution of the first nine surfaces can barely be

seen. The reason has already been mentioned before. The on-axis field also suffers from transverse chromatic aberration because the OARs of varying wavelengths are decentered in the image plane. In order to reveal the real transverse chromatic aberration, the contribution of OAR is removed in Fig. 4-17 using the eF'-lines as an example. It can be seen that the full-field displays are nearly symmetric now. The contribution from surface 11 is also tiny. The main contribution is coming from surface 10.

4.6.3 Anamorphic photographic lens

The anamorphic photographic lens [4-3] analyzed in this section has 28 surfaces including 2 freeform surfaces, 1 aspherical surface, 7 cylindrical surfaces and 18 spherical surfaces. The 3D layout of the system is shown in Fig. 4-18. The basic data is given in Table 4-5. The stop is located in the rear part of the system and set as a circular shape. Therefore the exit pupil is nearly symmetric in the tangential and sagittal cross sections. The system transfers the rectangular object into a squared image. The ratio between the sagittal and tangential magnifications is 2.

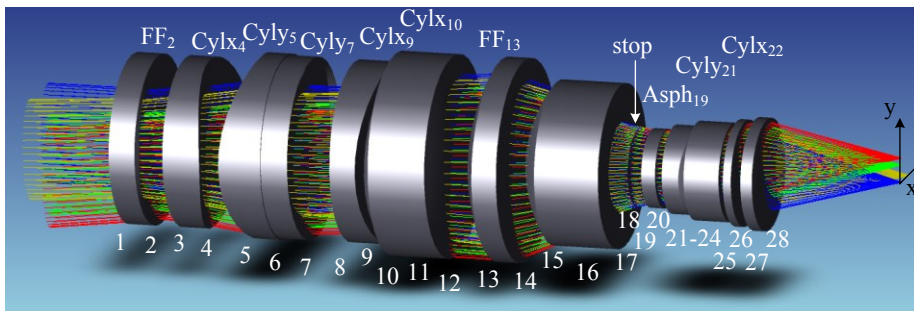


Fig. 4-18 3D layout of the anamorphic photographic lens.

| Parameters | Values |
|------------------------------|-------------|
| Wavelengths | F'eC'-lines |
| FOV $x \times y$ (°) | 48.4 × 21.4 |
| Tangential focal length (mm) | 50 |
| Sagittal focal length (mm) | 25 |
| Tangential F-number | 1.48 |
| Sagittal F-number | 1.31 |
| Effective F-number | 1.39 |

Table 4-5 Basic data of the anamorphic photographic lens. The effective F-number is the average of the tangential and sagittal F-numbers.

In order to evaluate the chromatic pupil aberration in such a complex optical system, the total chromatic aberrations of the outermost field corner $x=24.2^\circ$, $y=10.7^\circ$ are computed by both original and improved methods as shown in Table 4-6. The difference of 9.68% in the chromatic image variation and 4.55% in the y -component of transverse chromatic aberration

show that the chromatic pupil aberration in this system cannot be neglected. Therefore, the generalized chromatic aberrations in this system have to be calculated by the modified chief ray path.

| | Original | Improved | Error |
|--|------------------------|------------------------|-------|
| Chromatic image variation (F'C'-lines) | 11.992RL | 13.277RL | 9.68% |
| Transverse chromatic aberration x-component (eC'-lines) | 1.640D _{Airy} | 1.637D _{Airy} | 0.18% |
| Transverse chromatic aberration y-component (eC'-lines) | 0.189D _{Airy} | 0.198D _{Airy} | 4.55% |

Table 4-6 Comparison between original and modified calculations for the outermost field corner.

The full-field display of the chromatic image variation in the image plane is shown in Fig. 4-19. The chromatic image variations of the fields in the central region are positive. The chromatic image variations are negative for the fields at the edge. In between, two nodal points with zero chromatic image variation can be found on the x-axis with the field angles $\alpha = \pm 20.74^\circ$. The distribution of the full-field display is double plane-symmetric, which corresponds to the symmetry of the anamorphic system.

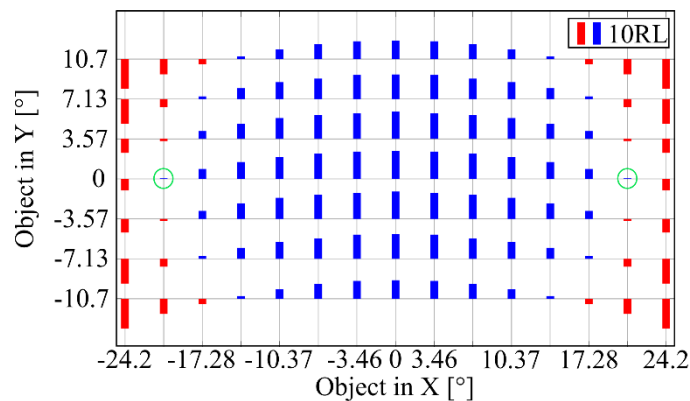


Fig. 4-19 Full-field display of the chromatic image variation in the image plane. Green circles indicate the nodal points without chromatic image variation. Red means negative, blue means positive values.

The full-field displays of the chromatic image variation contributed from the freeform surface 13 and the aspherical surface 19 are shown in Fig. 4-20. The bars indicate the difference between the on-axis field and other fields. It can be seen that the fields on the two axes have nearly the same contribution from the freeform surface. The irregular distribution for the off-axis fields in the four quadrants show the local slope change of the freeform. Compared with the freeform surface, the full-field distribution of the aspherical surface is much smoother because the aspherical surface is still rotationally symmetric.

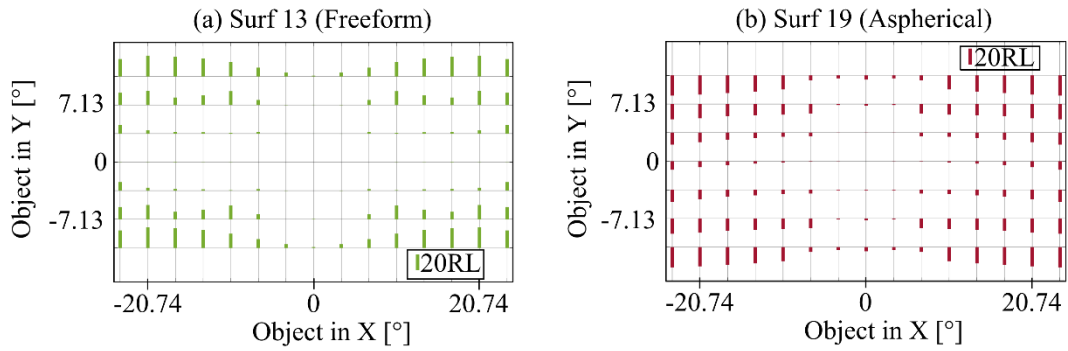


Fig. 4-20 Full-field displays of chromatic image variation on (a) freeform surface 13 and (b) aspherical surface 19 shown as the difference between the on-axis field and other fields.

The surface-by-surface chromatic image variations of the outermost field $x=24.2^\circ$, $y=10.7^\circ$ are shown in Fig. 4-21. Different types of surfaces are filled with different colors. The compensation of the chromatic image variation among the surfaces can be seen from the accumulated curve. Accordingly, the system can be classified into three lens groups (surfaces 1-9, 10-18, 19-28) in which the chromatic image variation is compensated. This general lens groups classification based on the achromatization matches the change of ray bundle diameter in the system.

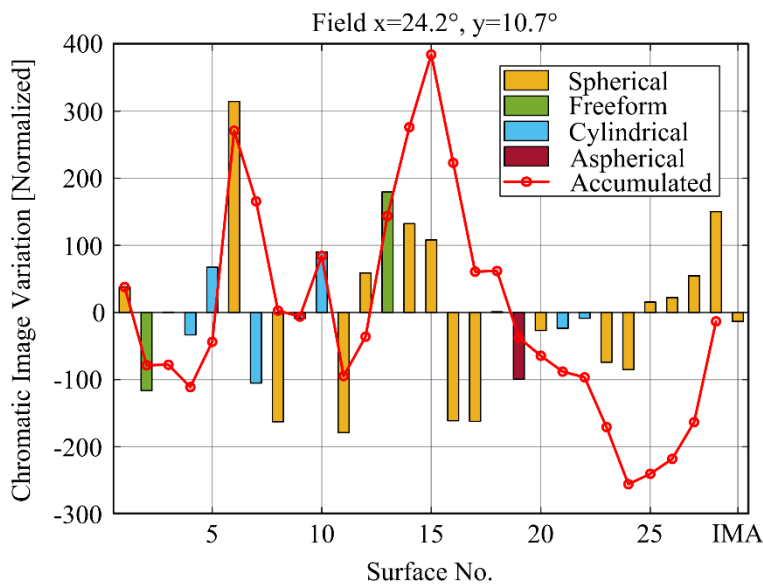


Fig. 4-21 Surface contributions to the chromatic image variation for the outermost field.

The full-field display of the transverse chromatic aberrations in the image plane is drawn in Fig. 4-22. The grid shows the intersection of the reference chief rays (e-line) in the image plane. The straight dashed lines demonstrate that the distortion in this system is well corrected for the whole FOV. The transverse chromatic aberrations between eF' -lines and eC' -lines are plotted individually as vectors pointing from the central wavelength to the boundary wavelengths. In rotationally symmetric systems, the two vectors are on the same line and point to the opposite direction because the incidence planes for all the fields do not

rotate from surface to surface. In non-rotationally symmetric systems, the two vectors are no longer on the same line and even pointing oppositely, which is demonstrated in this example.

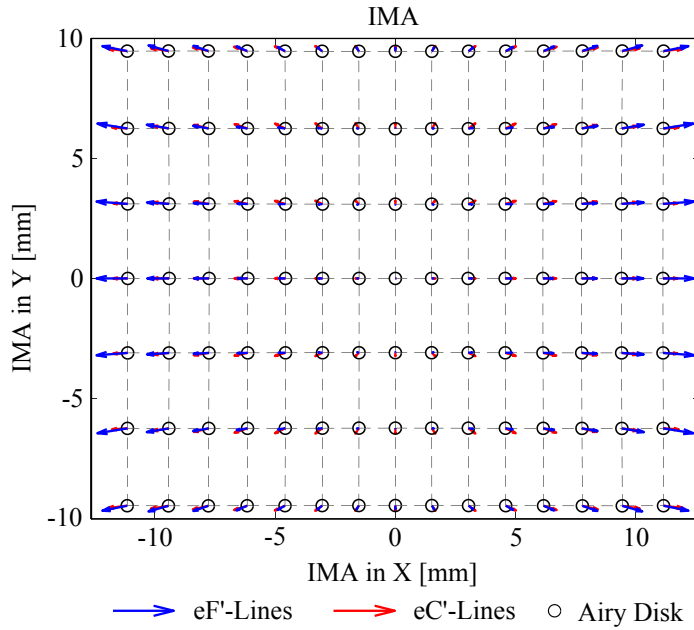


Fig. 4-22 Full-field display of the transverse chromatic aberration in the image plane.

It can also be observed that the fields in the central region have diffraction-limited size of transverse chromatic aberration and the transverse chromatic aberration of eF'-lines are smaller than eC'-lines. The situation in the edge region is more critical.

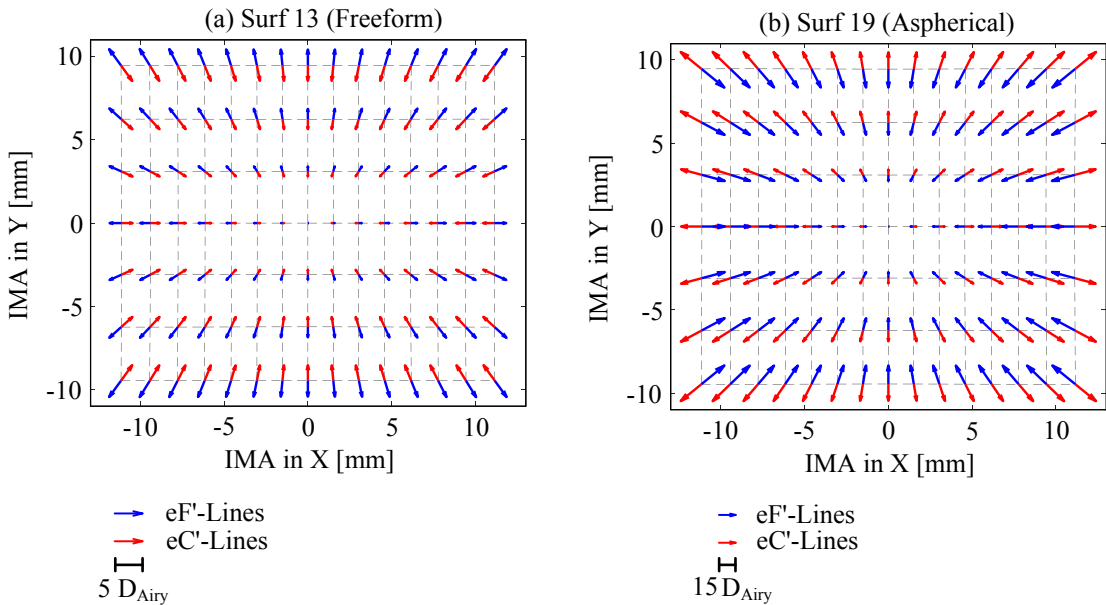


Fig. 4-23 Full-field displays of the transverse chromatic aberration on (a) freeform surface 13 and (b) aspherical surface 19.

Again, the full-field displays on the freeform surface 13 and aspherical surface 19 are plotted in Fig. 4-23. On the freeform surface, the extensions of the arrows never cross the center point (0, 0), which are different from the spherical surface because of the local slope change

on the freeform surface. The contributions on the two example surfaces are opposite to each other, but the magnitude on surface 19 is nearly three times larger than surface 13.

The surface contributions of the transverse chromatic aberrations in x - and y -directions between the eF^2 -lines are shown in Fig. 4-24 for the outermost field. In this anamorphic lens, both freeform surfaces have smaller contributions than the aspherical surface. The difference between the x - and y -directions on the freeform surface 2 is larger than the aspherical surface because the x - and y -radius of curvatures are decoupled on the freeform surface. However, the effect is not significant because the freeform surfaces have spherical basic shape. The decoupling effect is more obvious for the cylindrical surfaces in which the x - and y -radii are fully decoupled.

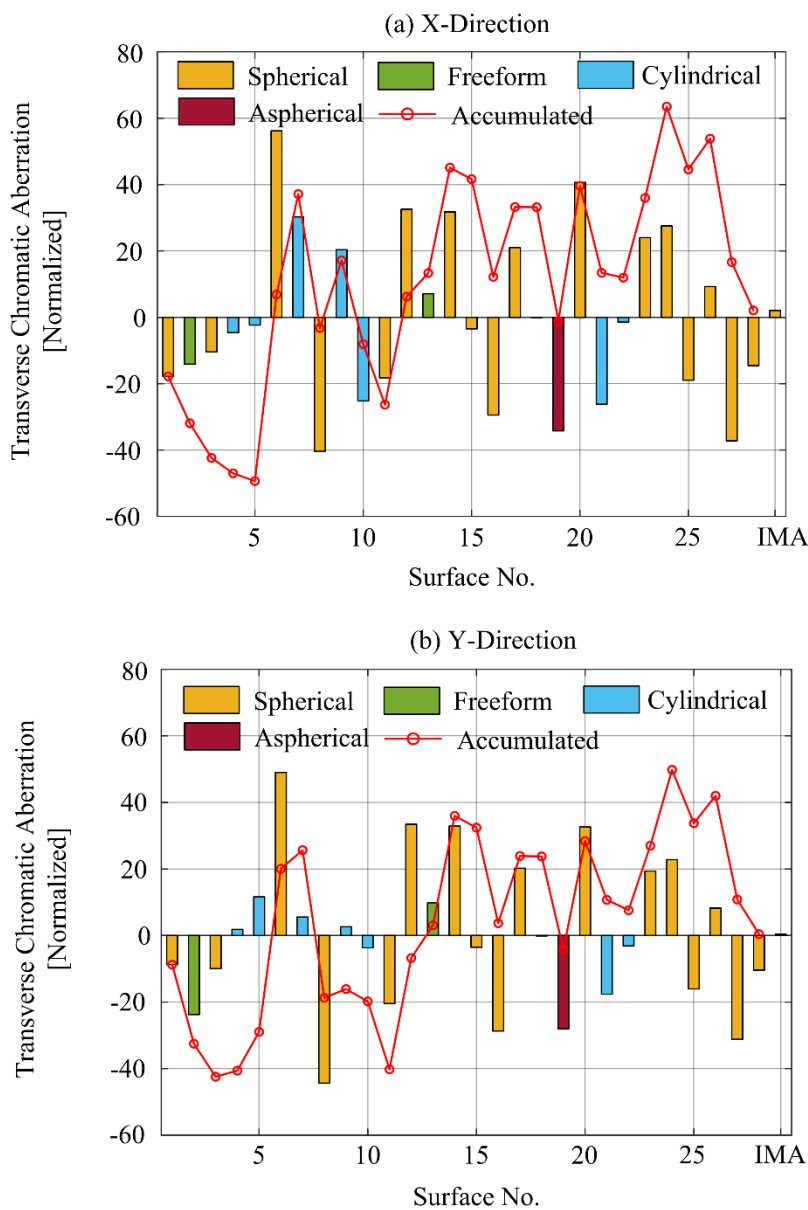


Fig. 4-24 Surface contributions to the transverse chromatic aberration between the eF^2 -lines for the outermost field corner in (a) x - and (b) y -direction.

4.6.4 HMD system

The performance of the generalized chromatic aberrations in HMD systems is of interest because this kind of system contains both refractive and reflective surfaces. The HMD system designed by Chen *et al.* [4-4] is employed here for analysis. The system layout is shown in Fig. 4-25 with a front stop. The first refractive surface is an anamorphic freeform; the second surface is a planar mirror; the third surface is an anamorphic freeform mirror; the last refractive surface is spherical. The basic data are listed in Table 4-7. Although the system was designed for a single wavelength, F'eC'-lines are used here to show the chromatic properties of the system.

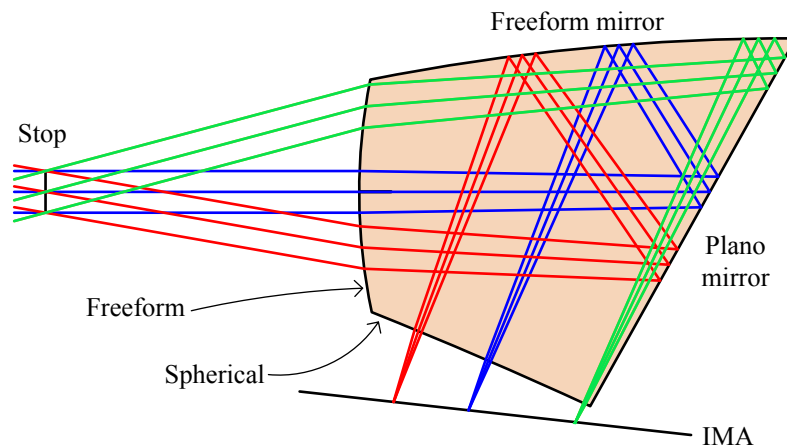


Fig. 4-25 Layout of the HMD system.

| Parameters | Values |
|----------------------|----------------------------|
| FOV ($x \times y$) | $50^\circ \times 30^\circ$ |
| Image-sided NA | 0.05 |
| Wavelengths | F'eC'-lines |

Table 4-7 Basic data of the HMD system.

The surface-resolved chromatic image variations of the on-axis field are plotted in Fig. 4-26. It can be seen that only the first and the last refractive surfaces contribute to the total chromatic image variation. The contributions of the inner two mirrors are zero. Both refractive surfaces have positive contributions because they both have positive focal power and are made of the same material. Therefore, there is no chance to correct chromatic image variation in this system.

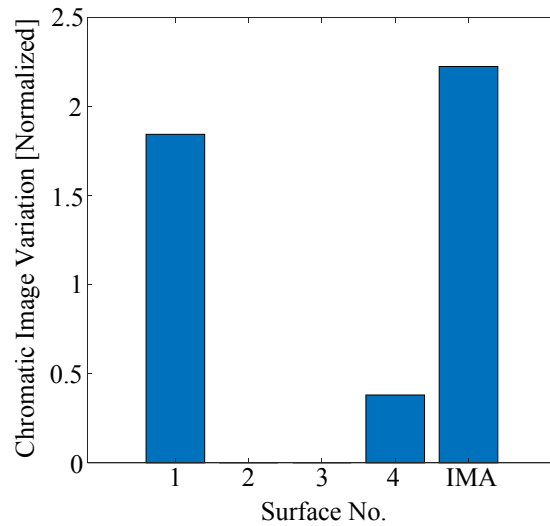


Fig. 4-26 Surface contributions to the chromatic image variation for the on-axis field.

The full-field displays of the chromatic image variation from the first surface to the final image plane are shown as the difference to the on-axis field in Fig. 4-27. The distributions are all plane-symmetric correspond to the structure of the HMD system. It is worth noting that the contributions for the fields of the two mirror surfaces are non-zero except the on-axis field. However, in common sense, the mirrors should not have chromatic aberrations. The violation is due to the induced chromatic aberration coming from the first refractive surface. The OAR is perpendicular to the first freeform surface, which means that the chromatic OARs do not separate after the first surface, and the ray paths for all the chromatic OARs are the same in between the surfaces. When the OARs are reflected by the mirrors, they have the same input parameters such as object distance, incidence angle and refractive index. Therefore, they have the same image distance, and the contributions of the mirrors are zero. However, the chief rays are inclined to the first refractive surface for the other fields, and the chromatic CRs are separated after refraction. They have different object distances when they reach the mirrors, resulting in the chromatic image variation. It is important to clarify here that the chromatic image variations generated at the mirrors are the induced aberrations from the former surface but not the intrinsic contribution of the mirrors.

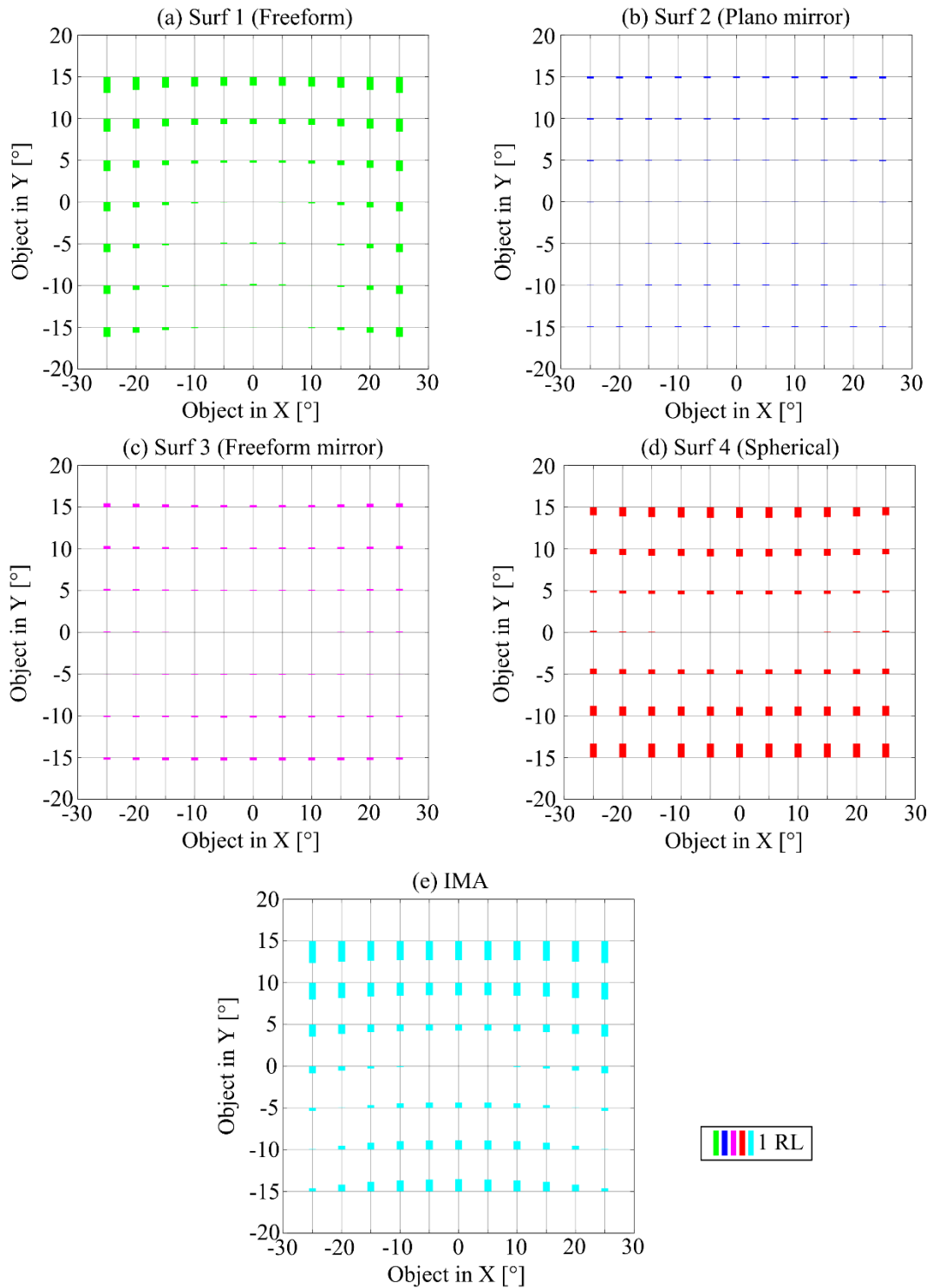


Fig. 4-27 Full-field display of the chromatic image variation plotted as the difference to the on-axis field on each surface (a)-(d) and in the image plane (e).

In Fig. 4-28, the surface contributions to the chromatic image variation are classified into the intrinsic and induced aberrations for the on-axis field and the outermost corner field. It can be clearly seen that the on-axis field has no induced aberration, but the outermost field has induced aberration starting from the second surface. The intrinsic aberrations on the mirrors are zero.

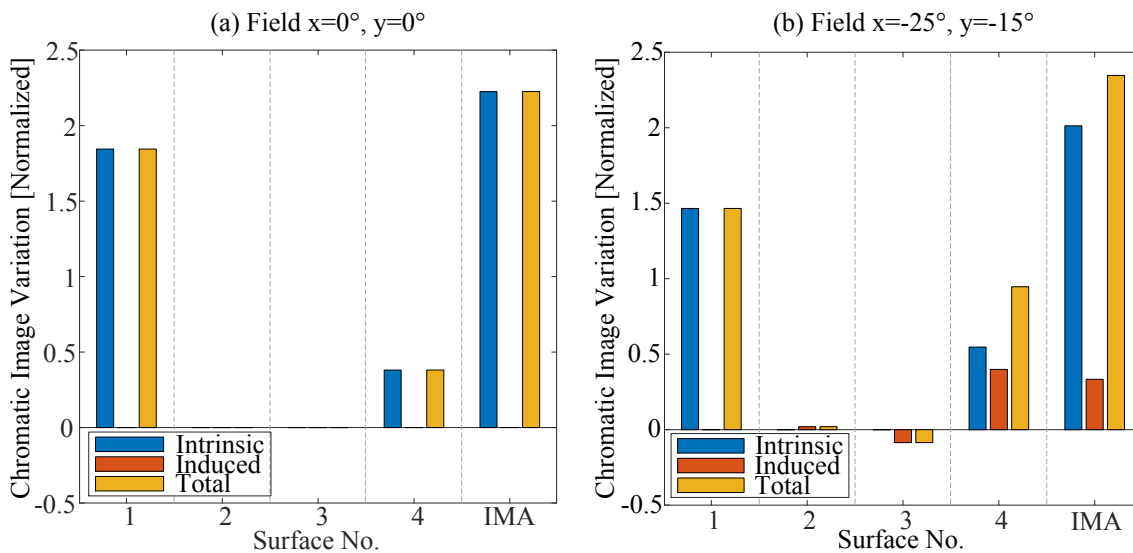


Fig. 4-28 Intrinsic, induced and total contributions on each surface for the (a) on-axis field and (b) outermost field.

The full-field transverse chromatic aberrations in the image plane are drawn in Fig. 4-29. The distorted grid of the central wavelength shows that distortion is not well corrected in the system. The Airy disk can hardly be seen in the plot, which shows that the extremely large transverse chromatic aberrations are remained for the whole FOV.

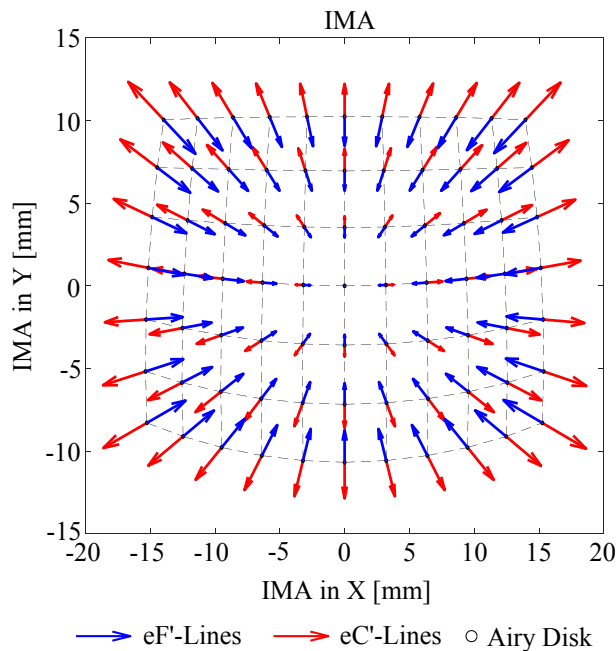


Fig. 4-29 Full-field display of transverse chromatic aberrations in the image plane.

The surface contributions to the transverse chromatic aberrations are shown in Fig. 4-30 in the full-field display. Obviously, for all the fields, the major contribution are coming from the first refractive freeform surface. The two mirrors have exactly zero contribution, which is different from the case of chromatic image variation. This is because the angular displacement between the chromatic CRs does not change if there is no material dispersion. Therefore, the mirrors have neither intrinsic nor induced transverse chromatic aberrations.

Also, the last refractive freeform surface has nearly no contribution. The reason can be checked from the system layout in which the incidence angles of the CRs are nearly zero at the last surface, thus the angular displacement between the chromatic CRs is considerably small.

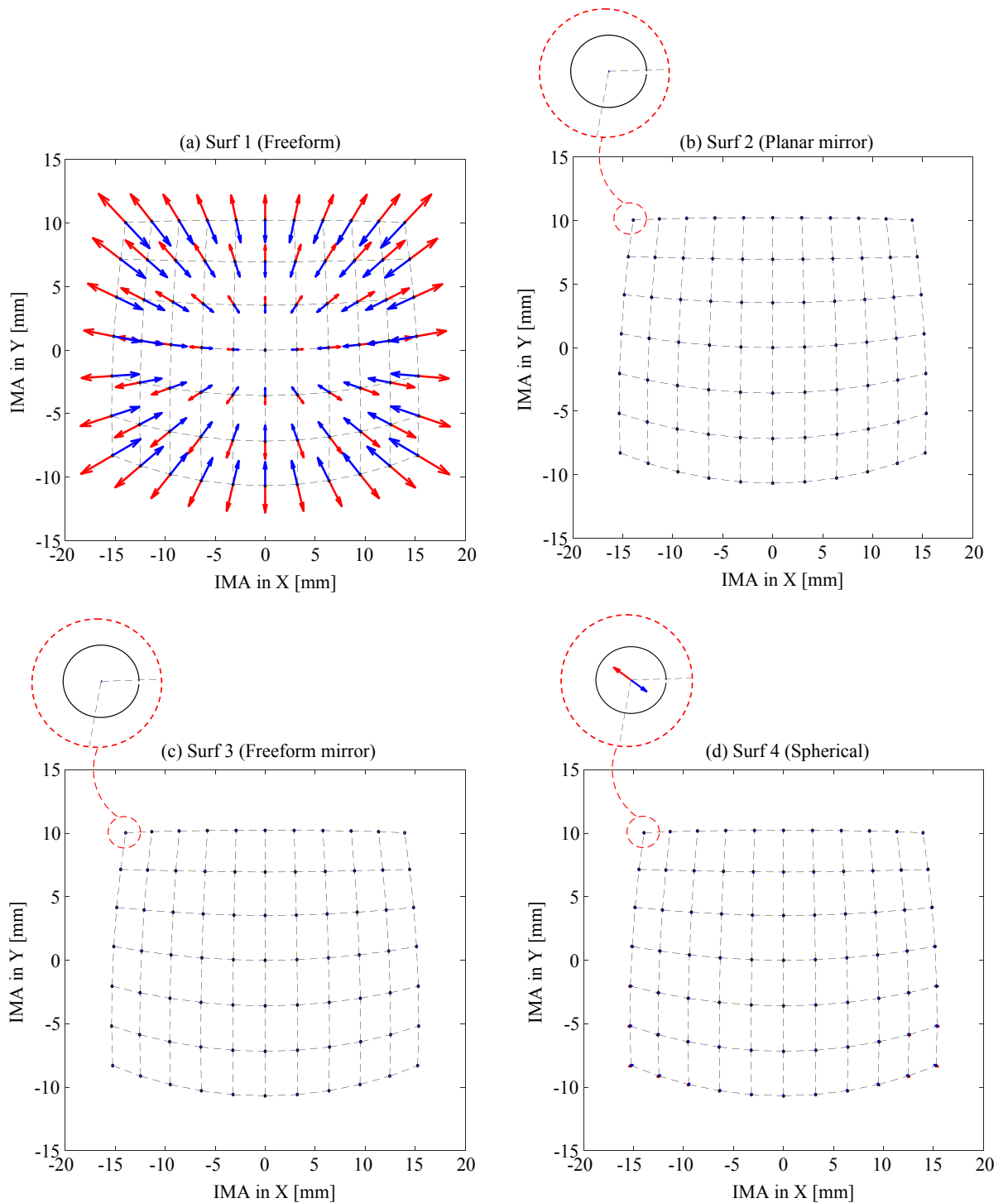


Fig. 4-30 Full-field transverse chromatic aberration drawn for each surface.

5 Tolerancing of Freeform System

Inspired by Sekman *et al.* who reconstructed the manufacturing surface errors by fitting the unwrapped phase pattern with polynomial expansions [5-1], a method to identify the boundary frequency between deterministic and statistical errors is put forward by understanding the behavior of the autocorrelation function of the unwrapped phase pattern in different frequency ranges. Then, a comprehensive model, which describes all types of surface errors, is established with the workflow chart. A synthetic surface is generated accordingly and inserted back to OpticStudio to analyze the influence of the surface errors and predict the performance of the real optical system.

5.1 Autocorrelation function of unwrapped phase pattern and its characterization

When investigating the manufactured freeform surface errors, the surface errors are transformed into the frequency domain by DFT. It takes only the amplitude part to plot the PSD. The phase information is lost. However, the phase terms determine the locations of the signals in the spatial domain. Sekman *et al.* unwrapped the phase function gained from DFT, which is between $\pm\pi$, by the transport of intensity equation (TIE) [5-2] to obtain a continuous phase pattern. According to their findings, the behavior of the unwrapped phase pattern between the deterministic and statistic errors is significantly different. This section continues with their work to benchmark the difference of the behavior by characterizing the autocorrelation function (ACF) of the unwrapped phase patterns in different frequency ranges. The ACF can be calculated by the Wiener - Khinchin theorem [5-3]

$$\text{ACF} = \text{real} \left(\text{iFFT} \left(\text{FFT}(\phi) \cdot \text{conj}(\text{FFT}(\phi)) \right) \right) \quad (5.1)$$

for discrete points. The phase pattern ϕ is normalized in advance. ACF shows the similarity between the data points. The values of ACF pattern are within ± 1 . 1 represents perfect correlation, -1 represents perfect anti-correlation, and 0 represents no correlation.

In the following, three examples – the first two belonging to deterministic errors and the last one belonging to statistical errors – are given to show the properties of their ACFs. Then, the difference between the ACFs is summarized.

As a first example, assuming a local surface error described by the RBF with an amplitude of 1mm, a center at (0.5, 1.5) and $w=0.5\text{mm}$. The surface is shown in Fig. 5-1 (a). Its 2D

PSD and unwrapped phase pattern in the frequency domain are shown in Fig. 5-1 (b) and (c). The FT of a Gaussian function remains Gaussian shape, but with the shape factor

$$w' = \frac{1}{\pi w} \tag{5.2}$$

The logarithmic PSD shows a clear boundary at $4w' = 2.55/\text{mm}$. 99.9937% of the total energy is inside the circle with a radius of $4w'$ [5-4]. It means that the phase inside this area is meaningful, but the phase outside is dominated by digital noise. This can also be seen in the phase pattern that the digital noise behaves smoothly with the same speed, and results in a tilted background plane. Therefore, it is important to distinguish between the meaningful and meaningless phase contributions.

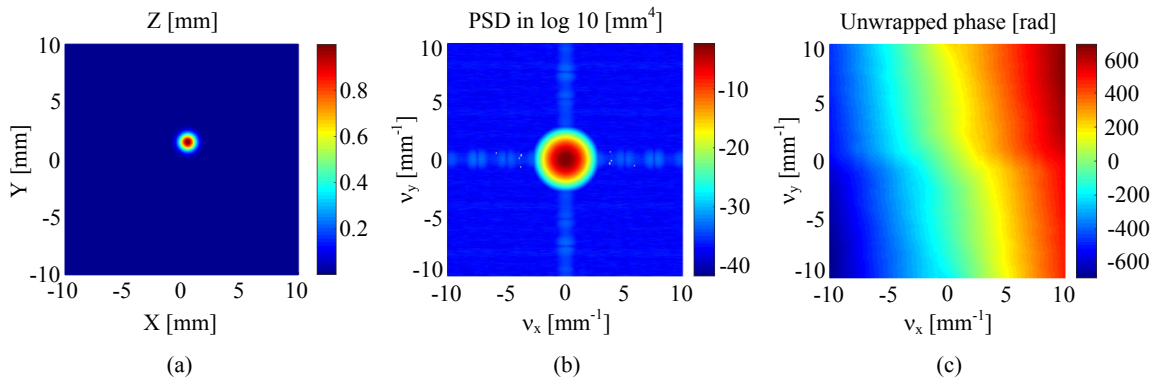


Fig. 5-1 (a) Surface error described by a Gaussian function in the spatial domain, and its (b) PSD and (c) unwrapped phase in the frequency domain.

Taking a square in the frequency domain with a half width of 4/mm so that the selected area is slightly larger than the meaningful area, then removing the background tilt in the unwrapped phase pattern, the PSD and unwrapped phase are plotted in Fig. 5-2. It can be noted in the phase pattern that the boundary between the meaningful and meaningless areas is sharp. The tilt of the meaningful pattern relates to the shift of the center of the Gaussian function in the spatial domain.

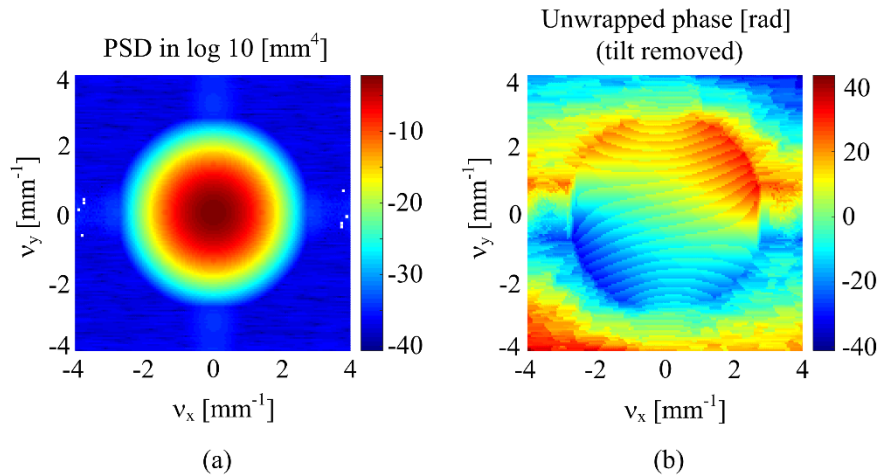


Fig. 5-2 (a) 2D PSD with the half width of 4/mm, which is a little bit large than $4w'$ and (b) the corresponding unwrapped phase pattern after removing tilt.

Then, applying several square windows with the half widths from 1 to 4/mm and equal spacing of 0.5/mm on the unwrapped phase pattern (hereafter, the background tilt in the unwrapped phase pattern is removed in default) to choose the sub-patterns in different frequency regions and calculate their ACFs. Fig. 5-3 shows the ACFs for seven different windows. In the first four frequency regions, the ACFs keep the same width. Then, for larger frequencies, the width of ACF shrinks. Therefore, the half width at half maximum (HWHM) of the central peak can be utilized to measure the weight of the deterministic error in a specific frequency region.

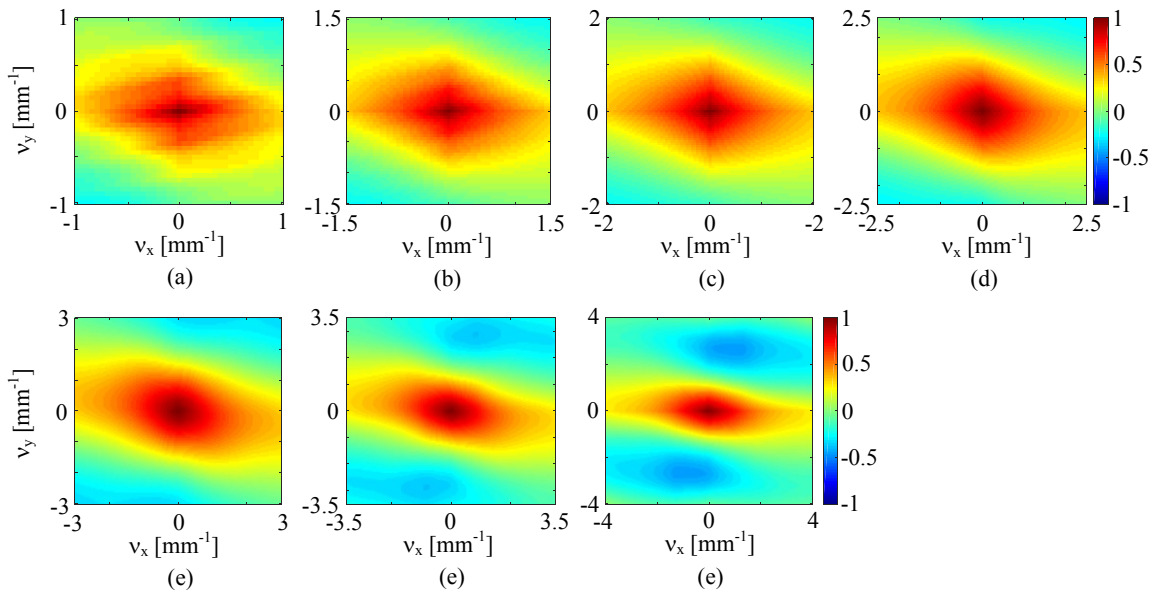


Fig. 5-3 ACF of the unwrapped phase pattern for seven frequency regions.

Since the ACF patterns are anisotropic, the average of the HWHMs in both x - and y -directions (hereafter denoted as HWHM) is computed and plotted in Fig. 5-4 (a) for each frequency region. The HWHM is linearly increasing within the meaningful region, which is shadowed. The dashed line shows the linear fit of the first four points. The last three points are clearly deviated to the dashed line, which reveal the influence of digital noise. In Fig. 5-4 (b), the normalized HWHM in each frequency region, which is the ratio between the HWHM and the maximum frequency in the corresponding region, are drawn. The normalized HWHMs within the meaningful region are nearly the same, again indicating that the HWHMs have the same slope. The last three normalized HWHMs are decreasing that correspond to the decrease of the HWHM in the ACF pattern.

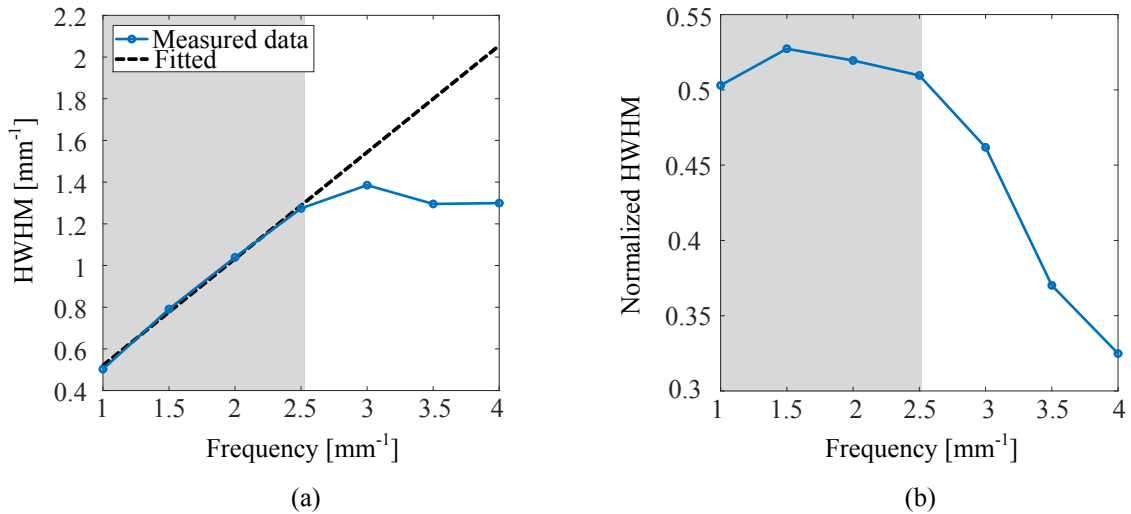


Fig. 5-4 (a) HWHM and (b) normalized HWHM in the ACF of unwrapped phase pattern. The frequency region in which the phase pattern contains meaningful signals is shadowed in grey.

The second example shows the cosine ripples with a period of 0.5mm to verify the property of ACF for the deterministic errors. The periodic structure, the 2D PSD with a selected boundary at 4/mm that is slightly larger than the frequency of the ripples at 2/mm, and the corresponding unwrapped phase pattern are shown in Fig. 5-5. A specific local feature is shown in the area around 2/mm of the unwrapped phase pattern.

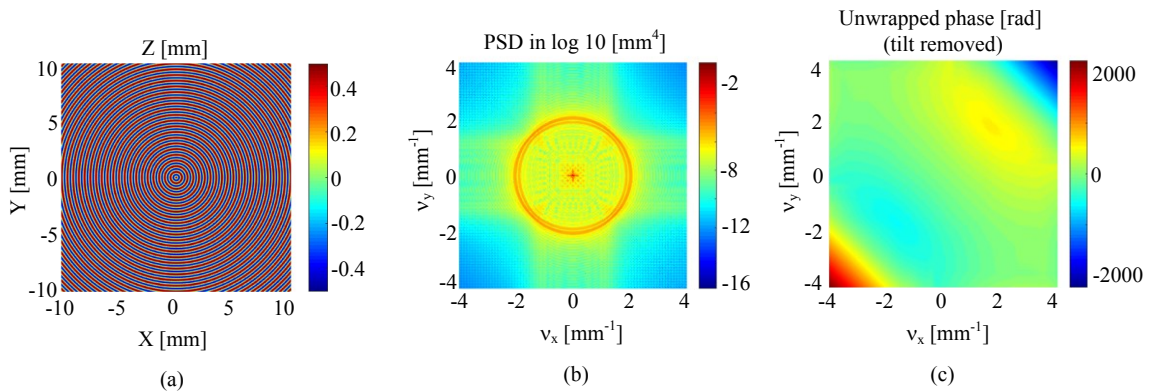


Fig. 5-5 (a) Cosine ripples with a period of 0.5mm in the spatial domain, and their (b) PSD and (c) unwrapped phase in the frequency domain.

Again, seven equally sampling windows with the same sizes as the cases in the Gaussian function are applied to the unwrapped phase to calculate the ACFs. The change of HWHM of the ACFs is drawn in Fig. 5-6 (a). It is interesting to see that, although the deterministic frequency is at 2/mm, the linear growth of the HWHM curve extends to the frequency of 2.5/mm. This is because the spike in the PSD has certain width as shown in Fig. 5-7 due to the finite number of rings in the region of interest. The normalized HWHMs are plotted in Fig. 5-6 (b), which has the same trend as the Gaussian function. The frequency region away from ripple frequency is dominated by digital noise.

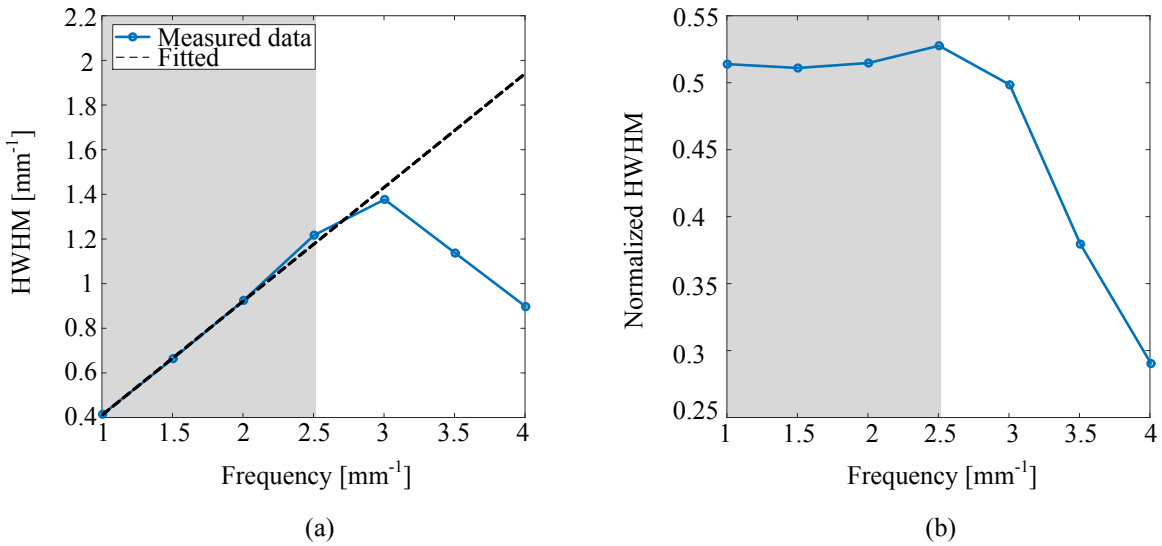


Fig. 5-6 (a) HWHM and (b) normalized HWHM in the ACF of unwrapped phase pattern. The frequency region in which the phase pattern contains meaningful signals is shadowed in grey.

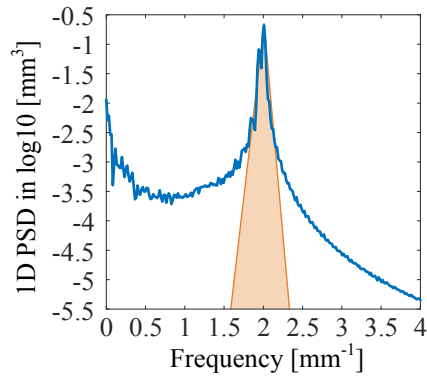


Fig. 5-7 1D PSD of the cosine ripples. The orange triangle shows the broadening of the spike at 2/mm.

In the third example, the statistical surface errors are generated with the direct method introduced in section 2.6.4 as shown in Fig. 5-8. The surface has a width of 1mm, which is smaller than the size of the surfaces described by the Gaussian function and cosine ripples. The maximum frequency for evaluation is selected as 40/mm, which covers the meaningful region in the center and the meaningless region near the edge. Since both the statistical surface errors and digital noise generate fluctuating unwrapped phase patterns, it is hard to distinguish a clear frequency boundary between them. However, it can be identified from Fig. 5-8 (c) that the local structure varies faster in the center for the statistical errors and slower in the edge for the digital noise.

Seven equal sampling windows from 10/mm to 40/mm are selected to plot the change of HWHM in the ACF of the unwrapped phase with increasing frequency region. The seven points are nearly located on the same line. There is a small oscillation at the right side indicating the small influence of digital noise. Compared to the linear fitted line of the cosine ripples extending to the frequency range of the statistical errors, the slope of the statistical

errors is significantly smaller than the slope of the deterministic errors. This also corresponds to the significantly smaller normalized HWHMs for the statistical errors.

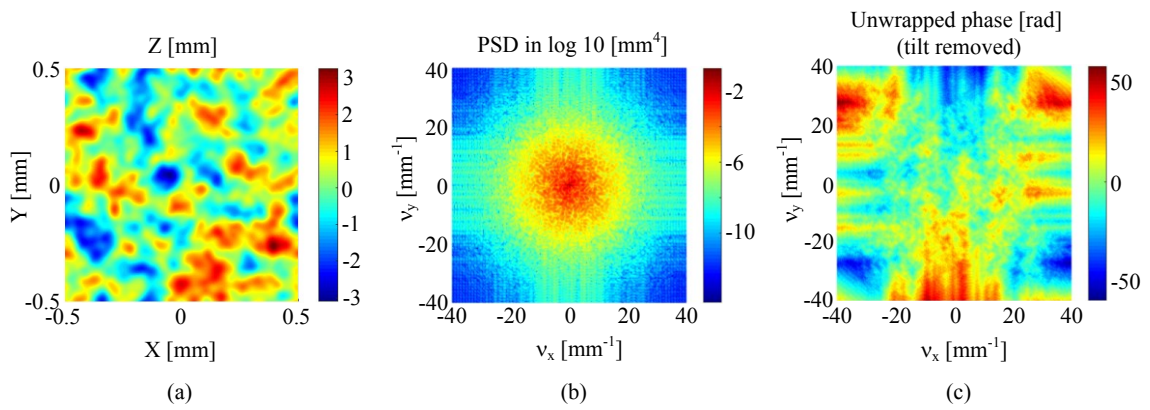


Fig. 5-8 (a) Statistical errors generated by the direct method in the spatial domain, and their (b) PSD and (c) unwrapped phase in the frequency domain.

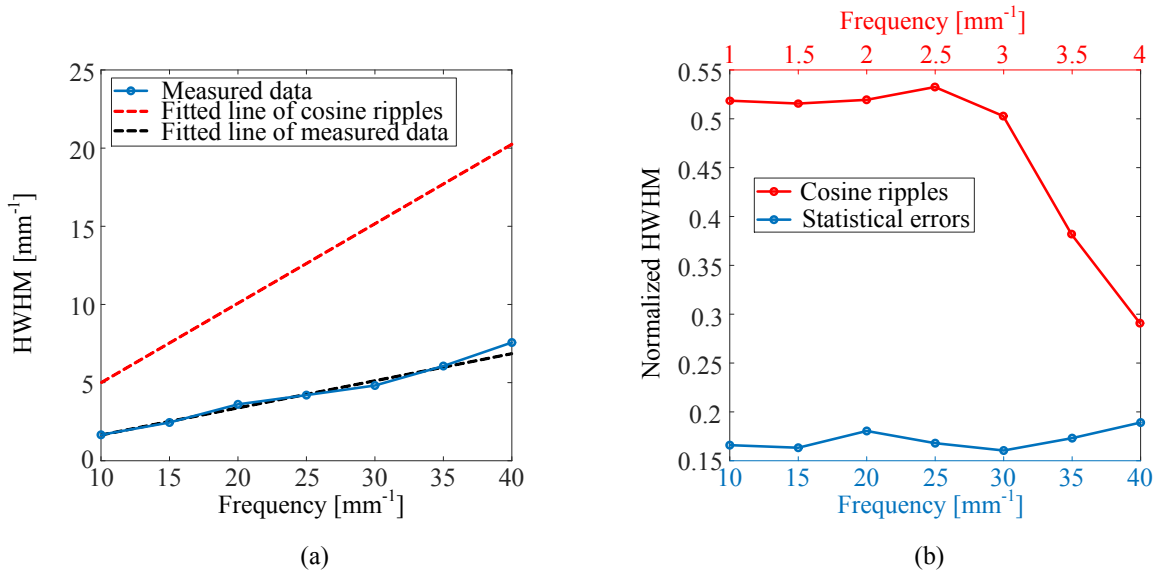


Fig. 5-9 (a) HWHM and (b) normalized HWHM in the ACF of unwrapped phase pattern.

It can be learned from the above examples that the deterministic errors cause specific continuous structures on the unwrapped phase pattern, the statistic errors arise strong local fluctuations. The widths of ACFs for these two kinds of errors are different. This can be viewed as a clear indicator to distinguish between deterministic and statistical frequency regions. However, the unwrapped phase pattern of the digital noise in the frequency range with extremely low amplitudes in the PSD should be excluded. The threshold of the meaningful and meaningless amplitudes is chosen to be about ten orders of magnitude smaller than the maximum amplitude.

5.2 Parametrization of real freeform surfaces

In practice, the surface errors across a broad spectrum are measured with white light interferometry (WLI) at different magnifications. The manufactured surface data examined in this section are provided by Fraunhofer Institute for Applied Optics and Precision Engineering IOF, Jena. The surface measurements inspect different sizes of the surface segment at four magnifications (1x, 2x, 10x, and 50x) to cover the spectral range from somewhere between LSF and MSF to HSF.

The manufactured surface errors of a diamond-turned surface are drawn in Fig. 5-10 with their 2D PSDs and unwrapped phase patterns. The maximum frequency selected for analyzing the phase spectrum at each magnification fulfills the principle discussed in the last section. It can be seen from the PSD patterns that the discrete deterministic areas in red are surrounded by the Fourier components of the statistical errors. Therefore, the unwrapped phase patterns look random because of the dominance of the statistical errors. The advantage of ACF computation is to reveal the difference between deterministic and random regions that is hidden in the noise-like pattern. Five equal sampling windows are applied on the unwrapped phase pattern at each magnification to calculate and show the trend of the ACFs in Fig. 5-11 (a). For the first three magnifications, the curves are linearly increasing in the beginning and then drop from the middle because the influence of statistical errors is getting stronger near the boundary frequencies. The curve is monotone increasing in the last magnification for the reason that the measured frequency range moves entirely to the HSF region, only the statistical errors can be seen. In Fig. 5-11 (b), the normalized HWHMs in the corresponding frequency windows are plotted. In addition, the averages of the normalized HWHMs at each magnification are connected in the black dashed curve. The average curve indicates that the influence of deterministic errors first increases and then decreases so that the boundary frequency between the deterministic and statistical errors lies between 40/mm and 60/mm.

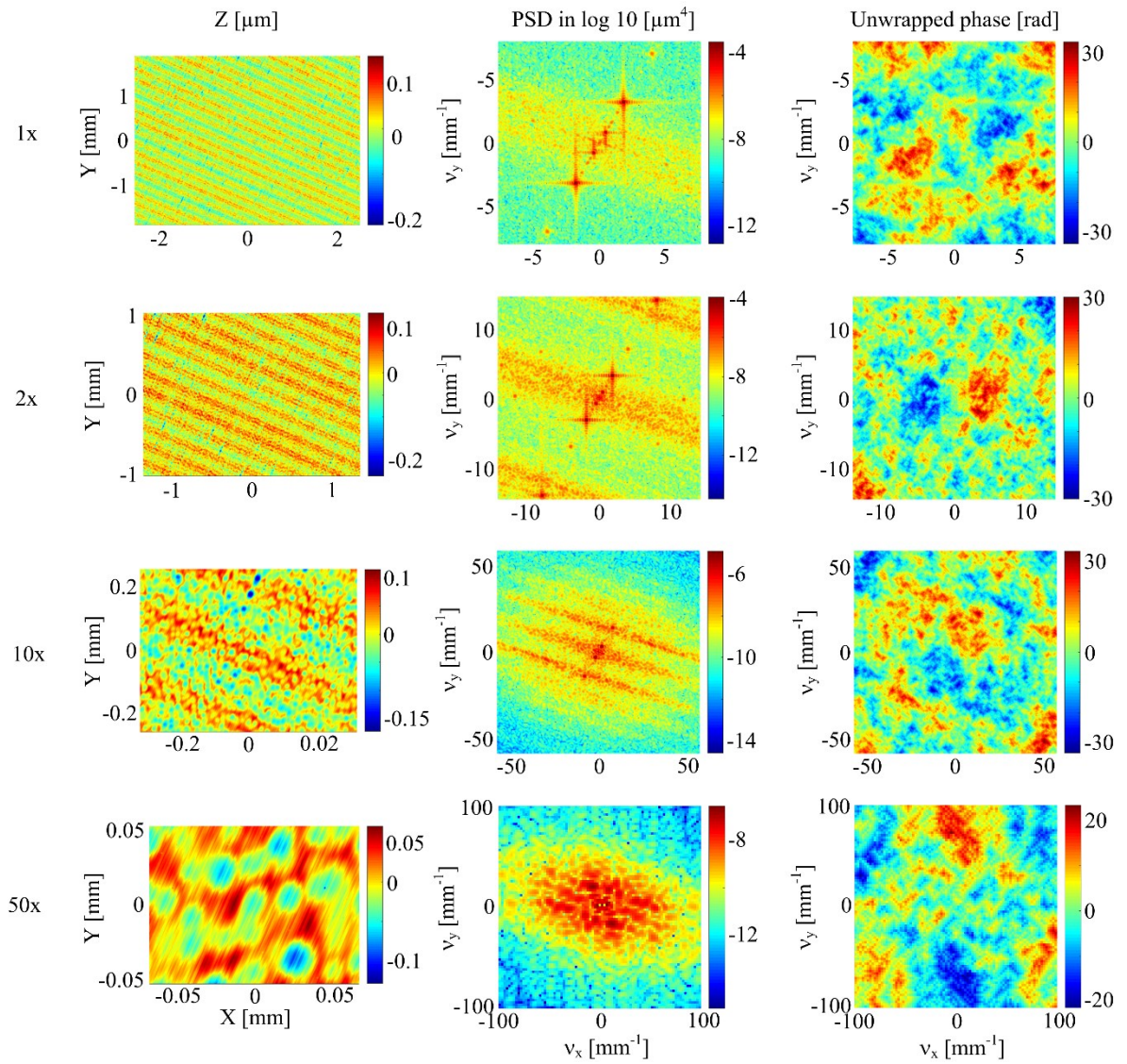


Fig. 5-10 Diamond-turned surface errors measured by WLI under four magnifications and their PSDs and unwrapped phases in the frequency domain.

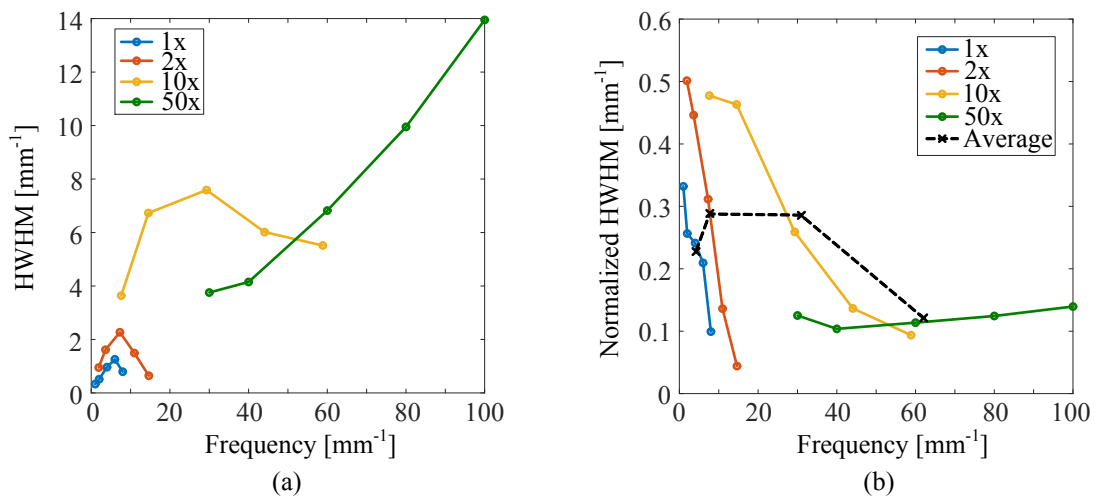


Fig. 5-11 (a) HWHM and (b) normalized HWHM in the ACF of unwrapped phase pattern for the diamond-turned surface measured at four magnifications.

The manufactured surface errors are measured again after polishing as shown in Fig. 5-12. Similarly, the ACFs of the unwrapped phase pattern with different window sizes are computed for each magnification. The HWHMs within different frequency regions are plotted in Fig. 5-13. It can be observed that the HWHMs at the magnification 10x are smaller than the corresponding values in the diamond-turned surface. In other words, the influence of the deterministic errors drops at this magnification. This can also be demonstrated from the average curve in the normalized HWHM diagram. Therefore, the polishing step removes the deterministic errors, which can be resolved under 10x magnification, and shifts the boundary frequency between the deterministic and statistical errors to the left, which is now between 20/mm and 40/mm.

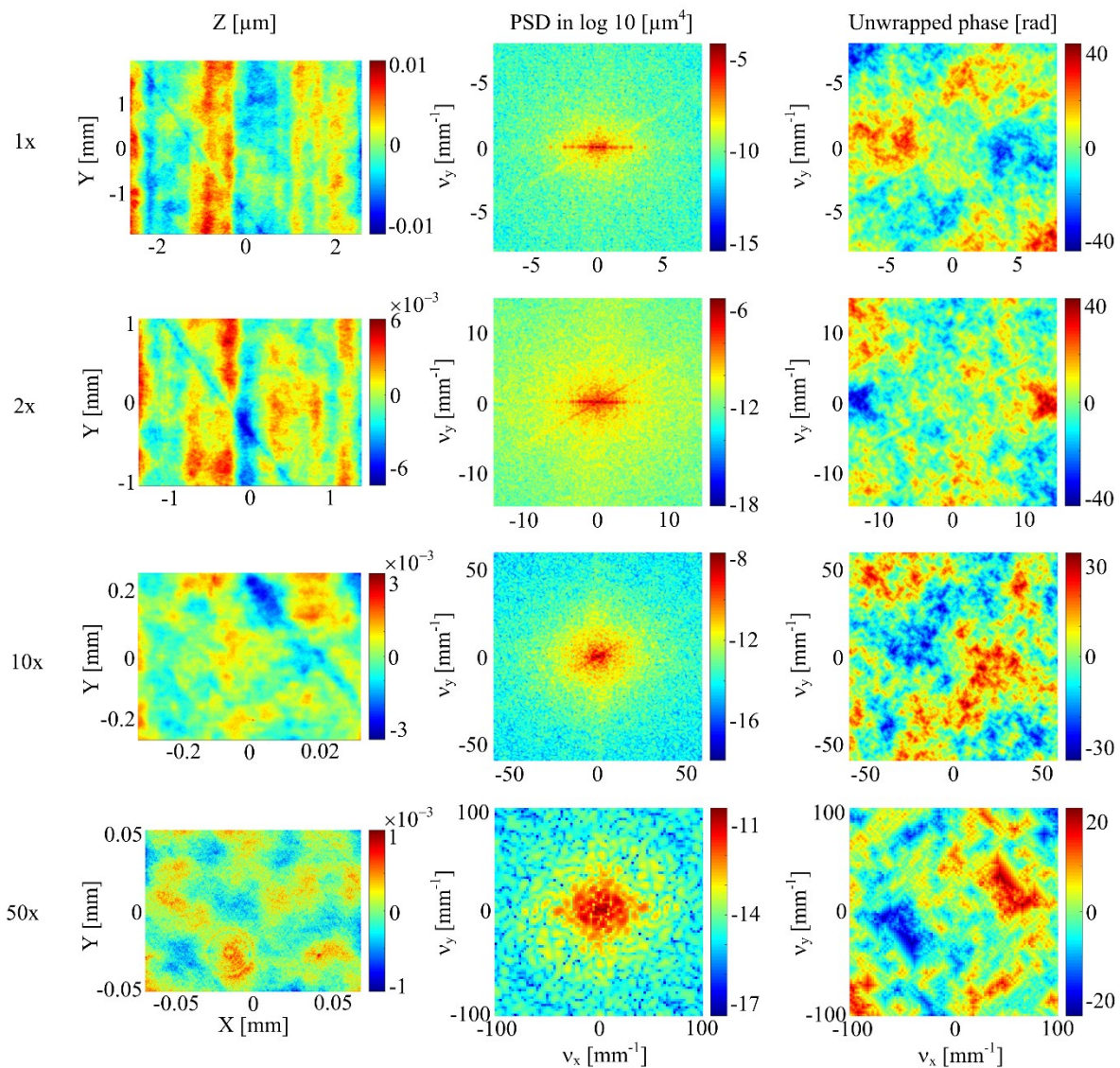


Fig. 5-12 Polished surface errors measured by WLI under four magnifications and their PSDs and unwrapped phases in the frequency domain.

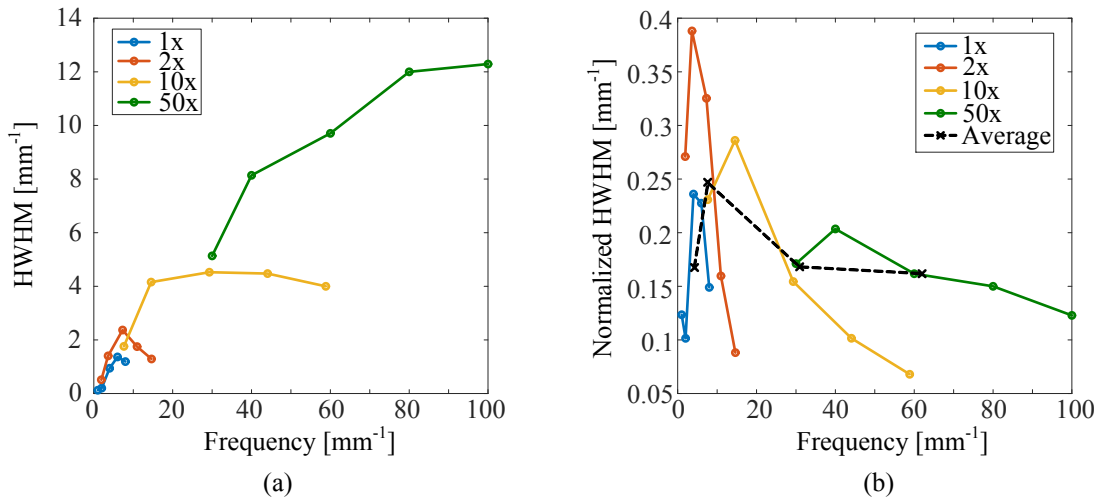


Fig. 5-13 (a) HWHM and (b) normalized HWHM in the ACF of unwrapped phase pattern for the polished surface measured at four magnifications.

To verify the boundary frequencies obtained by the characterization of ACF, 1D integrated PSD of each magnification and the average curve of the normalized HWHMs are plotted together in Fig. 5-14 for both diamond-turned and polished surfaces. In Fig. 5-14 (a), when the spikes are clearly shown in the PSDs, the values of the normalized HWHM are also high. In Fig. 5-14 (b), the spikes above 30/mm in the PSD of 10x are removed, the normalized HWHM is also reduced at this point.

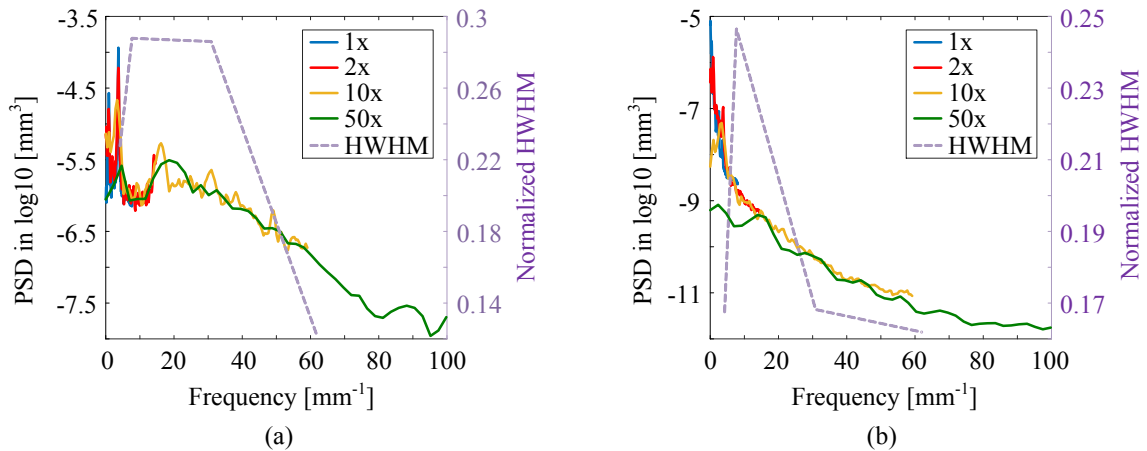


Fig. 5-14 Illustration of the relation between PSD and normalized HWHM for (a) the diamond-turned surface and (b) the polished surface.

5.3 Modelling a synthetic surface

Until now, the boundary frequency between the deterministic and statistical errors has been obtained by characterizing the ACF of the unwrapped phase pattern. The next step is to model the synthetic surface errors. The composition of the manufactured surface errors in the full spectral range is introduced in section 2.6 and can be written as

$$\Delta z(x, y) = \sum_j^{(Zernike)} a_j Z_j(x, y) + \sum_j^{(RBF)} b_j G_j(x, y) + \sum_j^{(ripples)} c_j R_j(x, y) + \sum_j^{(statistical)} d_j S_j(x, y) \quad (5.3)$$

where the Zernike polynomials describe the smooth change of the figure errors, RBFs describe the locally variant surface errors, then, the deterministic ripples and statistical errors are added.

The parameters that must be known before modelling are the RMS surface errors, boundary frequencies, and the frequencies and amplitudes of the ripples. In the beginning, the Zernike coefficients are tolerated according to the RMS value. Typically, up to 36 Zernike terms are chosen for tolerancing in order not to introduce strong local deviation [5-5]. Then, the sampling grid and the width of the RBF functions is determined, and the amplitudes of the RBF functions are generated randomly. The PSD of the LSF errors that are well described by summing up the Zernike polynomials and RBFs is calculated accordingly, and the slope of the linear part is fitted. The MSF errors can be divided into two parts. The PSD is extended into the MSF region to compute the MSF RMS error, which is the square root of the area under the PSD within this region. The MSF RMS error is used to generate statistical errors by the direct method. After identifying the envelope of the ripples, additional ripples are added using the functions listed in Table 2-5. The last step is to generate HSF statistical errors, which is identical to the process of generating MSF statistical errors. In this way, the full spectral surface errors are successfully modelled and then stored in the form of a “grid sag” surface required by OpticStudio.

The synthetic surface is reimported to OpticStudio to analyze the performance of a real optical system after manufacturing through raytracing. The image quality can be evaluated by spot diagram, PSF, and/or MTF. By performing the Monte Carlo simulation, the whole process can be repeated many times to predict the range of system performance. The workflow of the freeform surface modeling and system tolerancing can be visualized in Fig. 5-15.

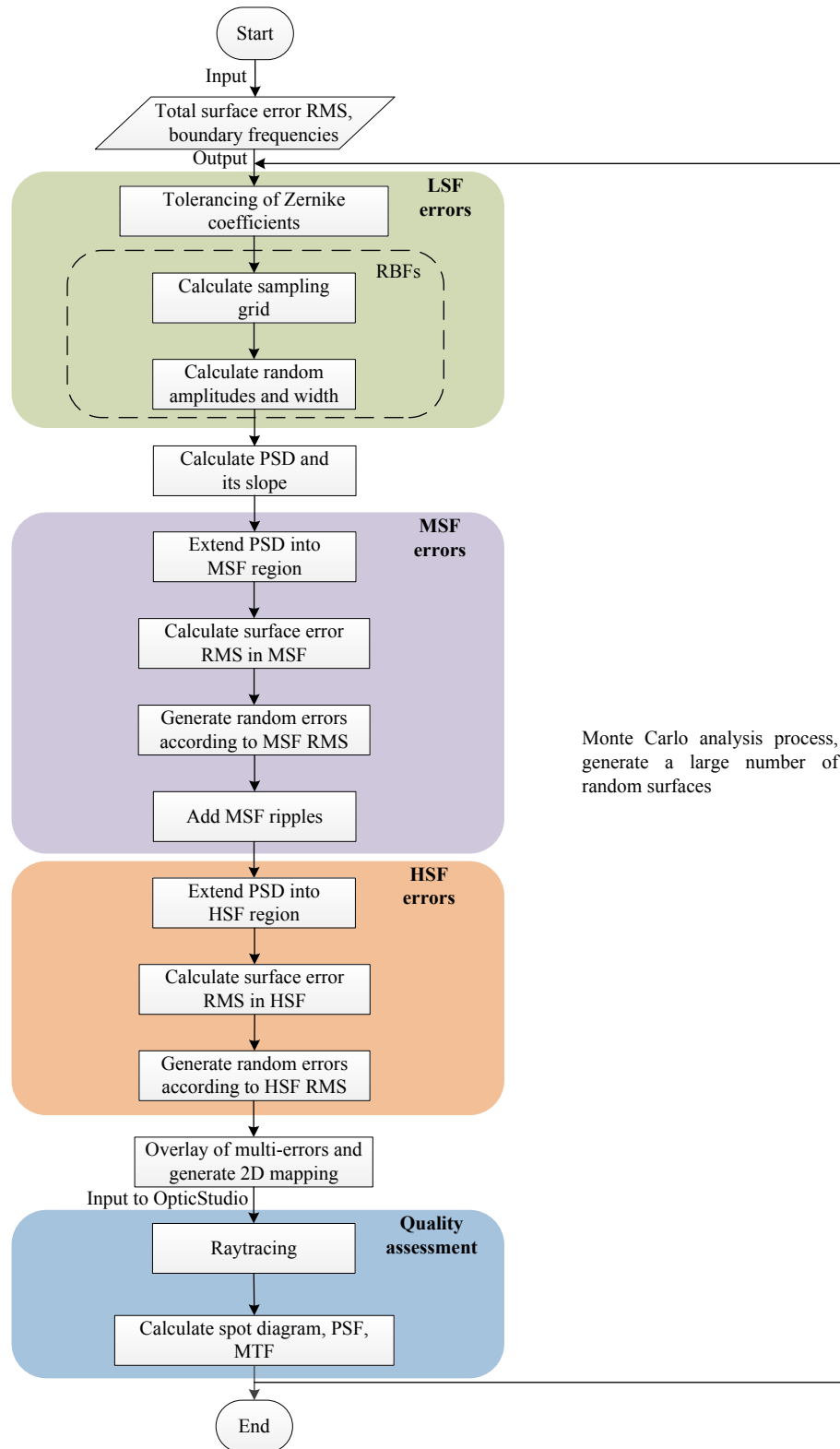


Fig. 5-15 Workflow of the freeform surface modeling and system tolerancing.

5.4 Example of tolerancing a freeform system

In this section, the HMD system introduced in section 4.6.4, which contains two freeform surfaces, is utilized again to show the tolerancing of a freeform system. As shown in Fig. 5-16, only the first refractive freeform surface is tolerated.

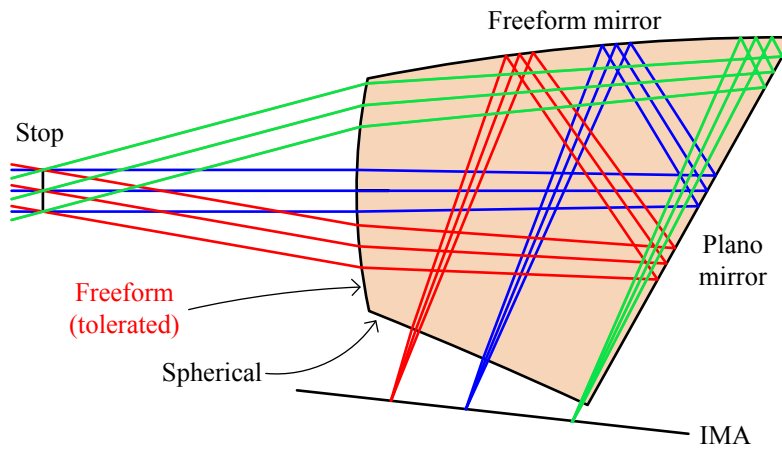


Fig. 5-16 The first freeform surface in the HMD system is tolerated as an example.

In the following, several simplifications and assumptions are clarified:

1. It is assumed to be a diamond-turned freeform surface with a total RMS surface error of 150nm, which is an empirical value given by practical experience [5-6].
2. The local deviations described by RBFs are of no concern here because the determination of the width of the Gaussian functions is not trivial and needs further study.
3. The surface has a rectangular aperture with $D_x=36\text{mm}$ and $D_y=24\text{mm}$. Considering the computation ability of the computer (Inter i7-4710MQ, 2.50 GHz, 16 GB memory), the sampling grid on the surface is fixed to be $n_x \times n_y = 3600 \times 2400$. Then, the highest frequency is obtained as

$$v_{x,\max} = v_{y,\max} = 50/\text{mm} . \quad (5.4)$$

The resolution in the frequency domain for both directions are

$$\Delta v_x = 0.028/\text{mm}, \quad \Delta v_y = 0.042/\text{mm} , \quad (5.5)$$

4. It has been concluded that the boundary frequency between the MSF and HSF regions is between 40/mm and 60/mm for the diamond-turned surface in the last section. However, the maximum frequency that can be resolved in the example surface is at 50/mm. Looking back to the PSD of the diamond-turned surface in Fig. 5-14, the two highest spikes are at 0.9/mm and 3.7/mm, the amplitudes of other spikes in the higher frequency region are much smaller. For the purpose of taking the full spectral errors into account within 50/mm, the spikes with low amplitudes are neglected. Then, it can be assumed here that the LSF, MSF, HSF regions are 0-1/mm, 1-10/mm, and 10-50/mm, respectively.

5. The PVs of the ripples are assumed to be 150nm. The centers of the ripples are located at the center of the surface for simplification.

The first 36 Zernike fringe polynomial terms except for the piston and tilt terms that can be removed by proper alignment are tolerated according to the total RMS surface error. The obtained figure errors are shown in Fig. 5-17 (a). The 1D integrated PSD of the figure errors are plotted in Fig. 5-17 (b). The linear part on the right is fitted by a line with a slope equal to -1.985.

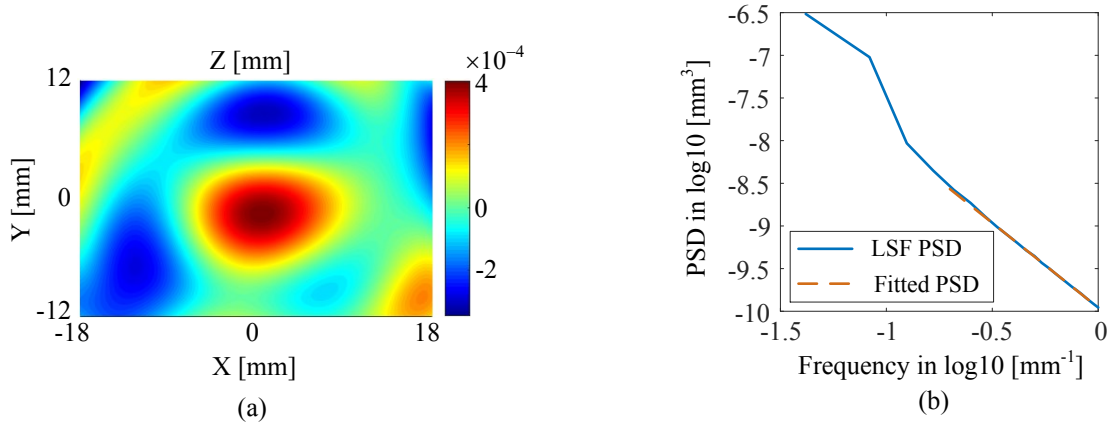


Fig. 5-17 (a) Figure errors and (b) its 1D integrated PSD.

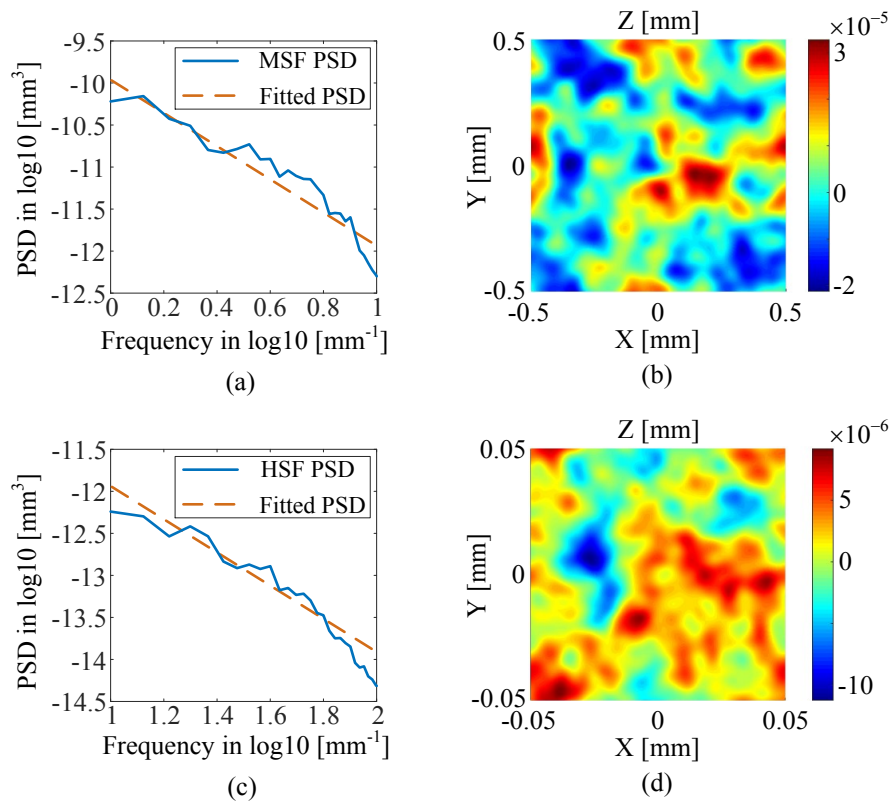


Fig. 5-18 Extension of the linearly fitted PSD into (a) MSF and (c) HSF regions. Statistical surface errors in (b) MSF and (d) HSF regions are generated accordingly by the direct method.

To simulate random errors in the MSF and HSF regions, the linearly fitted PSD is extended into both regions as shown in Fig. 5-18 (a) and (c). The area under the dashed line is the square of $\sigma_{\text{rms, rand}}$. By choosing proper slope factor k in the direct method (for both MSF and HSF random errors here, $k=0.3$), the random errors can be generated. The blue curves show the PSDs of the simulated errors, which coincide with the fitted lines. In Fig. 5-18 (b) and (d), the MSF and HSF random errors are shown in the surface segment $1 \times 1 \text{mm}^2$ and $0.1 \times 0.1 \text{mm}^2$, respectively. The MSF and HSF random errors in the whole surface are the repeat of the surface segments to fill the entire aperture. Fig. 5-19 shows the cosine ripples across the whole surface with the frequencies at $0.9/\text{mm}$ and $3.7/\text{mm}$, amplitudes of 150nm .

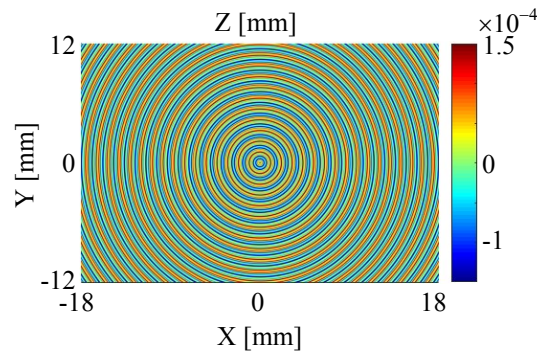


Fig. 5-19 MSF ripples.

Above all, three types of surface errors – LSF figure errors, ripples, and random errors – are fully defined. Their individual impacts on the system performance are investigated in the next step. The spot diagrams in the image plane of the central field ($0^\circ, 0^\circ$) and the maximum field on the x -axis ($25^\circ, 0^\circ$) are shown in Fig. 5-20 for the nominal system before tolerancing and the systems perturbed by individual surface errors. The changes of the RMS spot radius for these two fields from the nominal system to the system with LSF errors are 0.6% and 4.4%, respectively. The small impact of LSF errors can be understood from the ray bundle distribution on the freeform surface. The system layout in Fig. 5-16 shows that HMD system has a front stop with a large distance, which strongly separates the ray bundles on the first freeform surface. Moreover, the original freeform surface is described by an anamorphic asphere function until the 4th order, which has no local variation. Comparing the slopes of the original surface and the slopes of LSF errors in Fig. 5-21, the slopes of the LSF errors is 3 orders of magnitude smaller than the original surface. The circles indicate the ray bundle of each field on the surface. It can be seen that the change of surface slope due to LSF errors within a ray bundle can be neglected. For each field, the LSF errors can be viewed as an overlay of a tilted plane on the original surface and result in distortion. For example, for the central field ($0^\circ, 0^\circ$) after adding the LSF errors, the change of RMS spot radius is $0.03 \mu\text{m}$, but the change of chief ray intersection point in the y -direction is $1.4 \mu\text{m}$, which is more

significant. The ripples result in the broadening of spots with sharp boundaries. The random errors broaden the spots with irregular boundaries.

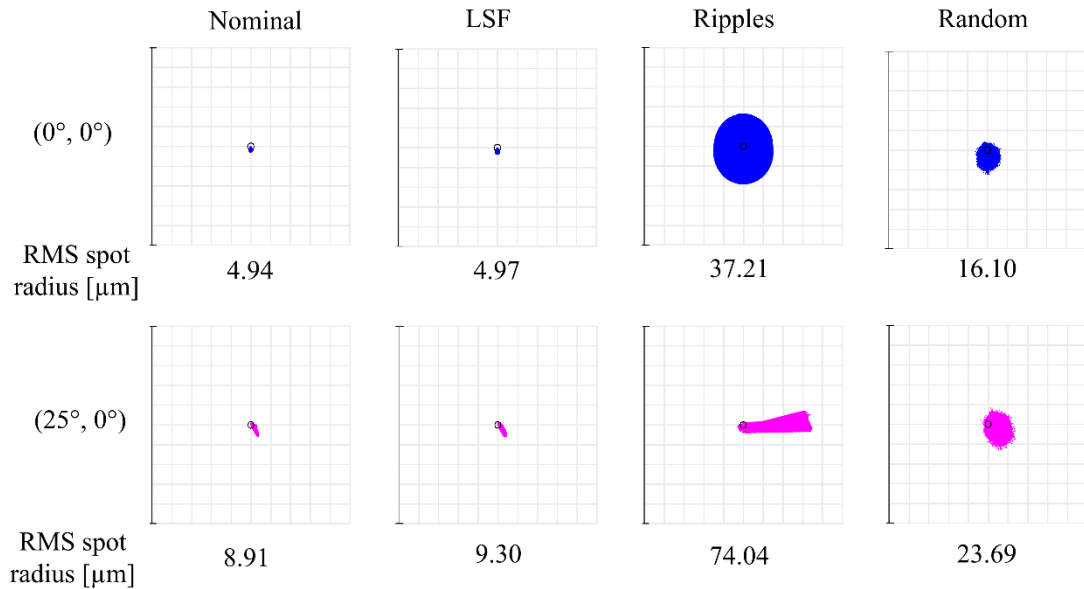


Fig. 5-20 Spot diagrams of the fields (0°, 0°), (25°, 0°) in the nominal system and perturbed systems with individual surface errors. The width of each spot diagram is 400μm.

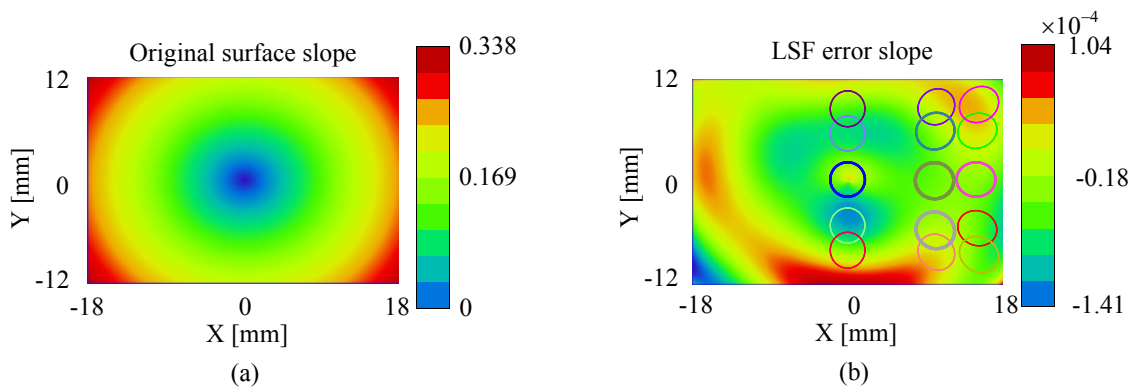


Fig. 5-21 Slope of (a) original surface and (b) LSF errors. The circles in (b) show the range of ray bundle from individual fields on the surface. The fields are sampled with the normalized coordinates 0, 0.7, 1 on both x - and y -axes.

The PSFs of the nominal system and perturbed systems with individual errors are shown in Fig. 5-22. Compared to the spot diagram, PSF shows the energy distribution in the image plane. The influence of the LSF errors is hard to be seen. The periodic structure of the ripples causes the diffraction effect, and thus, the neighboring diffractive orders can be clearly seen. In addition, the location of the central peak of the field (25°, 0°) is shifted due to non-zero incidence. The scattering effect introduced by the random errors is isotropic. The cross sections of the PSFs along the x -axis for the central field are plotted in Fig. 5-23 to show the impact of surface errors on the energy distribution. The small-angle scattering effect of the MSF ripples and the wide-angle scattering effect of the random error can be seen clearly.

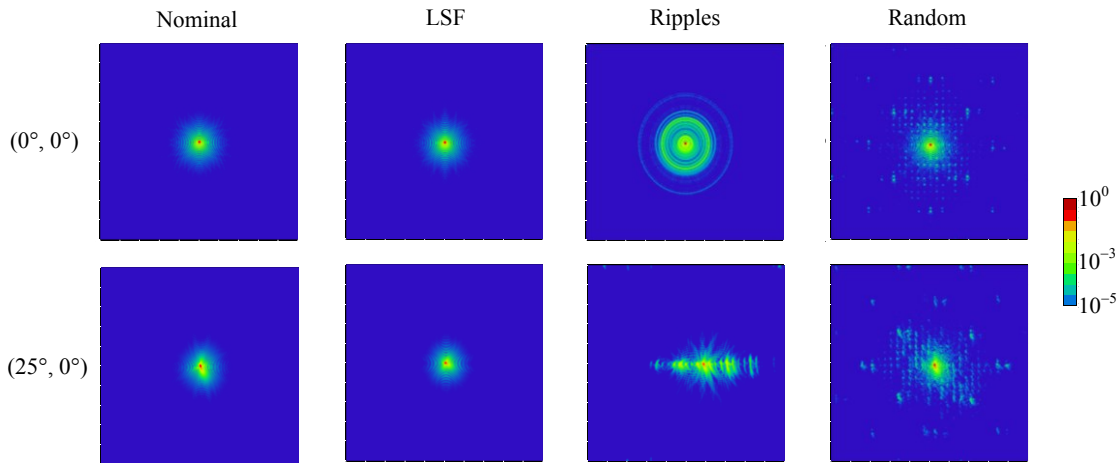


Fig. 5-22 PSFs of the fields (0°, 0°), (25°, 0°) in the nominal system and perturbed systems.

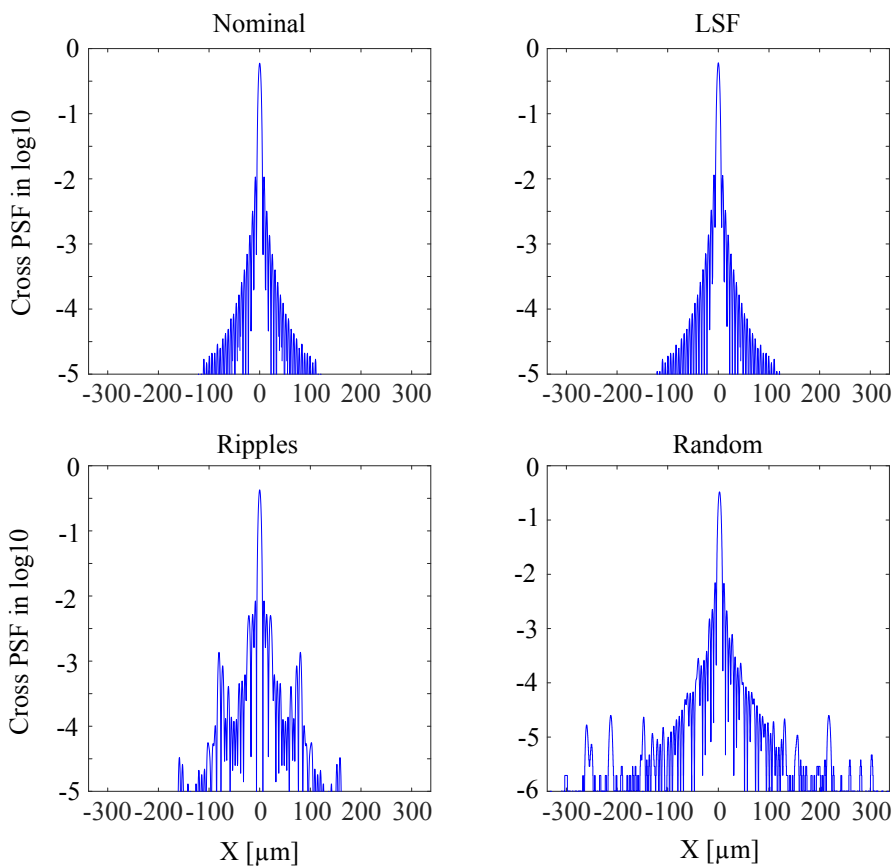


Fig. 5-23 Cross sections of PSFs along the x-axis for the central field in the nominal system and perturbed systems.

The MTFs of the nominal system and perturbed systems with individual errors are shown in Fig. 5-24, which allows us to analyze the system response in the frequency domain. The highest frequency in the MTF is set according to the maximum resolved frequency of the synthetic surface at 50/mm. The drops of MTF due to LSF and random errors are small. The drop of MTF due to wide-angle scattering is depending on $\frac{\sigma_{rms,rand}}{\lambda}$, in which λ is the working wavelength of the system. For this system in the visible spectrum, the drop of MTF can be

neglected. The waviness structure on the MTF of ripple errors is associated with the periods of the ripples. It can be seen that the tangential and sagittal MTF curves of the central field are oscillating with nearly the same period. The slight difference is coming from the quasi-circular exit pupil of the HMD system. For the field $(25^\circ, 0^\circ)$ on the x-axis, only the sagittal MTF curve oscillates because the ripples are only in the sagittal direction within the aperture of this field. The peaks of the waviness on the MTF are not decreasing smoothly because of the modulation of the ripples with two frequencies.

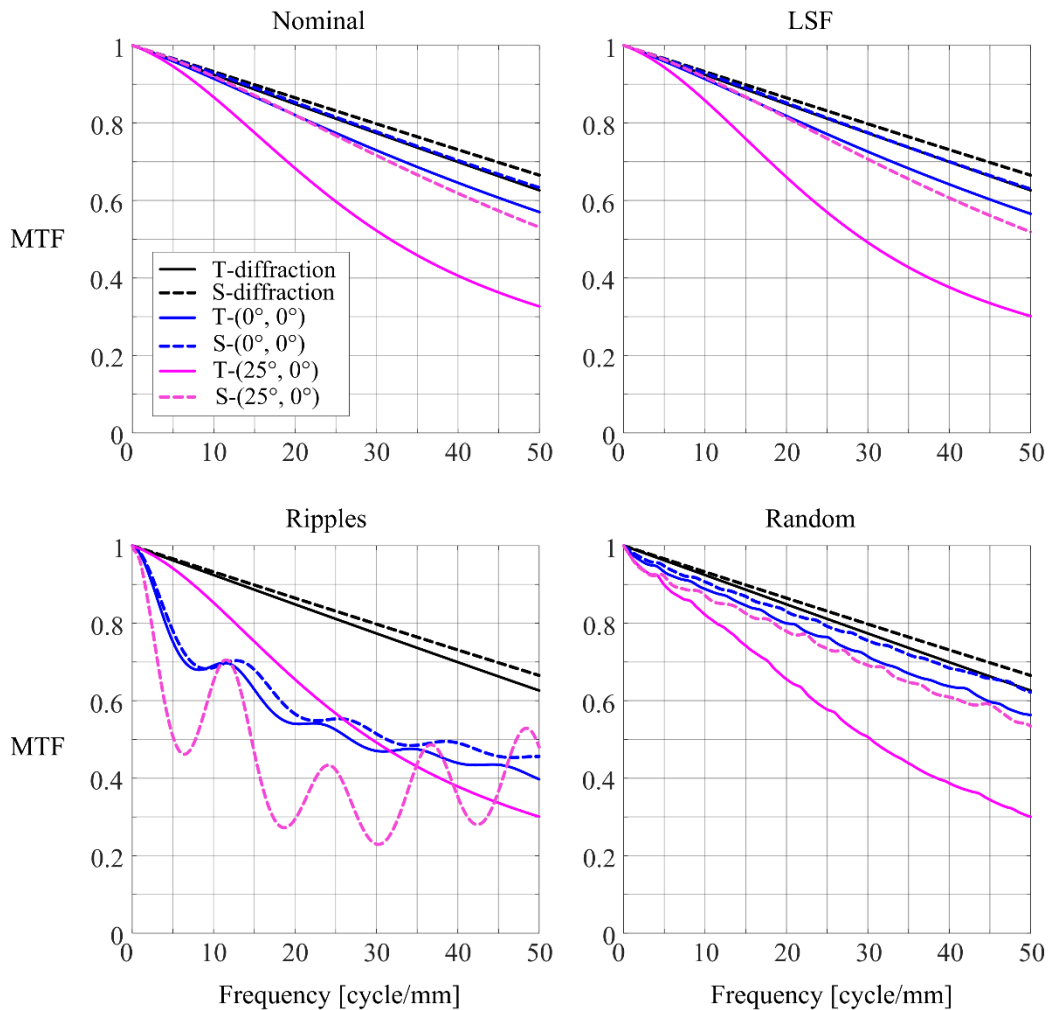


Fig. 5-24 MTFs of the fields $(0^\circ, 0^\circ)$, $(25^\circ, 0^\circ)$ in the nominal system and perturbed systems. T is the tangential cross section, S is the sagittal cross section.

The synthetic surface error in the full spectral range is the summation of the above individual surface errors. Fig. 5-25 shows the whole surface errors and two surface segments with the width of 1mm and 0.1mm to inspect the MSF and HSF errors. The lower right side draws the PSDs of the three surface sizes. The connection of the three PSDs shows the linear behavior and two deterministic spikes.

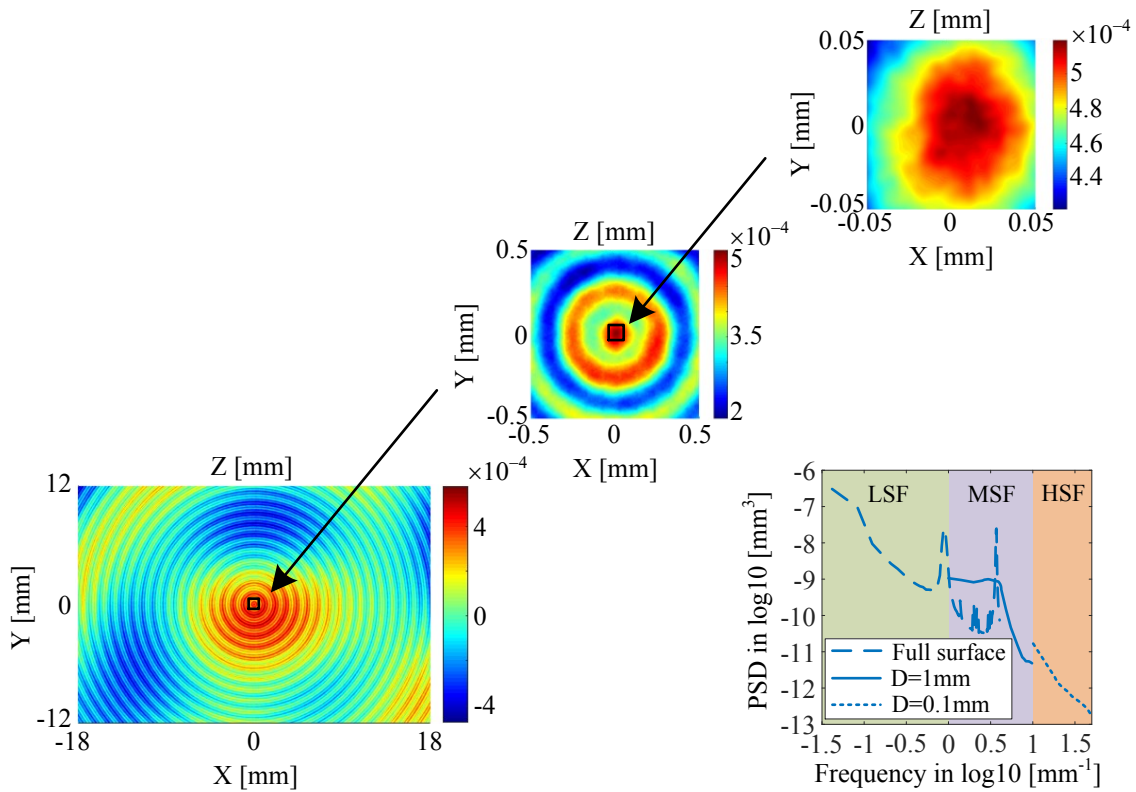


Fig. 5-25 Total synthetic surface errors with zoom-in inspection. The PSD across the whole frequency range is shown in the lower right side.

The spot diagram, PSF and its cross section, and MTF of the fields $(0^\circ, 0^\circ)$ and $(25^\circ, 0^\circ)$ are shown from Fig. 5-26 to Fig. 5-29. In all the plots, the impact of deterministic ripples is dominant. The effect of wide-angle scattering can be seen from the irregular boundary in the spot diagram and PSF.

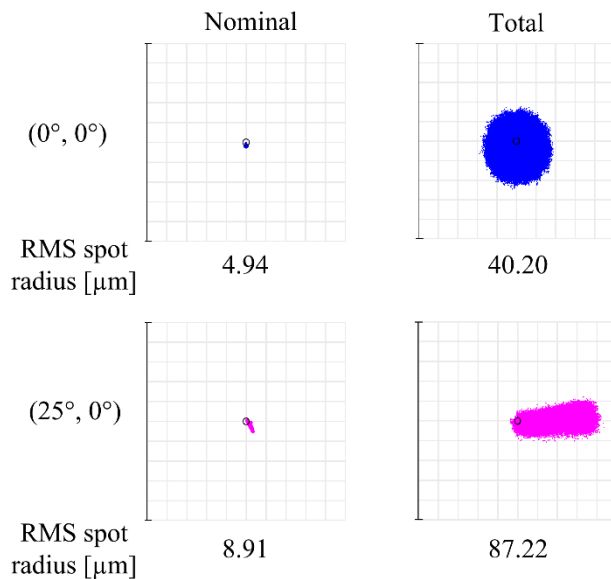


Fig. 5-26 Spot diagrams of the fields $(0^\circ, 0^\circ)$, $(25^\circ, 0^\circ)$ in the nominal system and the system with total surface errors.

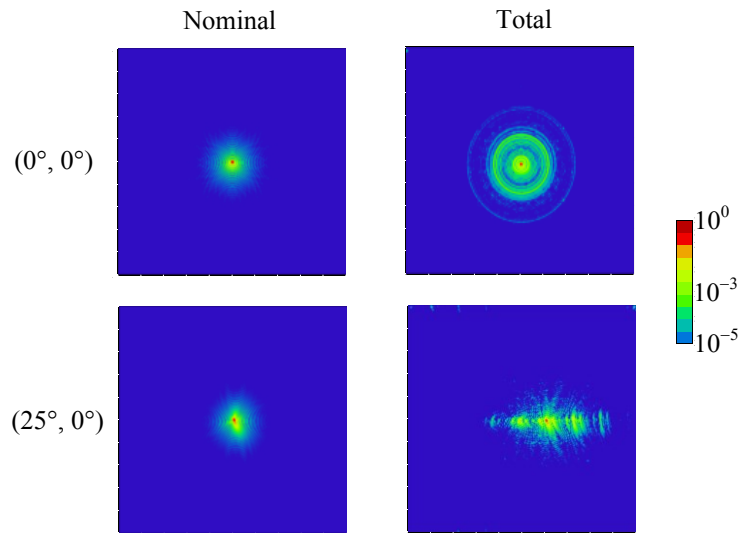


Fig. 5-27 PSFs of the fields $(0^\circ, 0^\circ)$, $(25^\circ, 0^\circ)$ in the nominal system and the system with total surface errors.

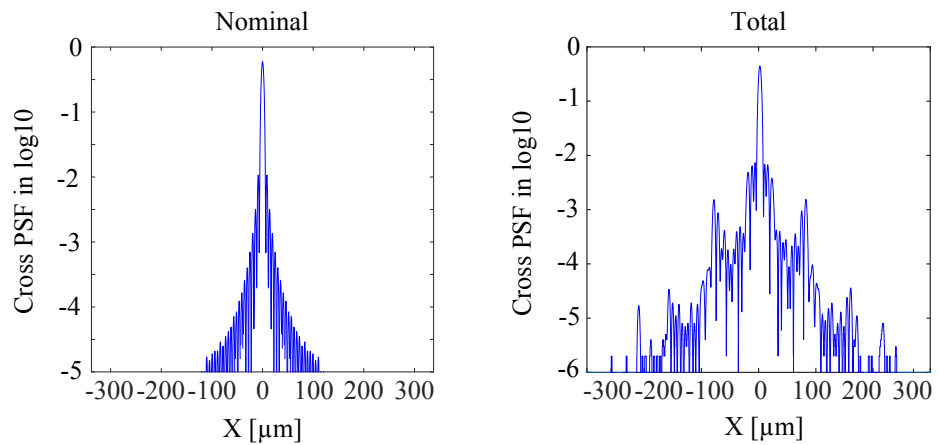


Fig. 5-28 Cross sections of PSFs along the x -axis for the central field in the nominal system and the system with total surface errors.

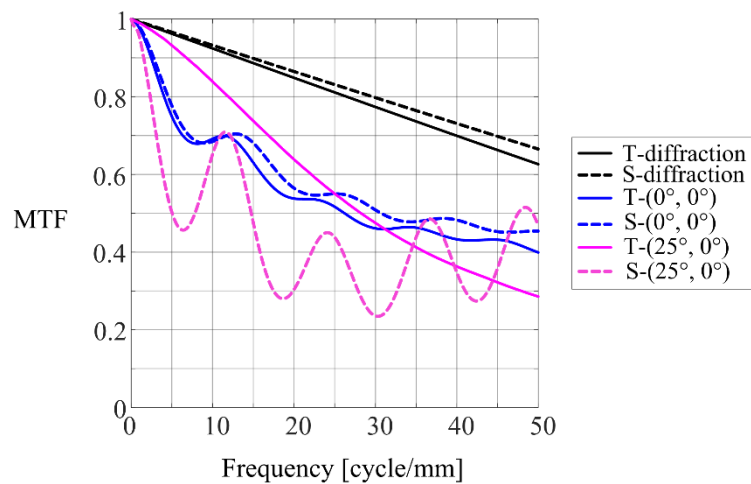


Fig. 5-29 MTFs of the fields $(0^\circ, 0^\circ)$, $(25^\circ, 0^\circ)$ in the nominal system and the system with total surface errors.

6 Conclusions

To improve the optical system development by introducing freeform surface and coordinate breaks, this thesis presents the generalization of obscuration elimination, chromatic aberration description, and tolerancing from rotationally symmetric optical systems to non-rotationally symmetric optical systems.

The obscuration issue that arises in reflective optical systems could be avoided by tilting the mirrors against the optical axis. In this work, four sub-error functions are proposed to describe all the obscuration situations in 3D reflective optical systems. Then, the four sub-error functions are multiplied by individual weighting factors, and then summed up to constitute a total obscuration error function to indicate the degree of obscuration in the system. By setting tilting angles of the mirrors as variables and selecting a proper optimization algorithm, an obscuration-free structure can be obtained without human involvement as the first example shows. In the second example, the ability of the method for a Yolo telescope to correct astigmatism is demonstrated. An initial structure is reached by optimizing both the obscuration error function and the astigmatism error function to be minimal. Then, by introducing freeform surfaces without perturbing the mirror position, an unobscured Yolo telescope with large FOV and satisfactory image quality is achieved.

The design of non-rotationally symmetric refractive systems gives rise to the demand on the evaluation of the chromatic aberrations. This thesis defines generalized chromatic aberrations by setting up clear reference axis and plane and distinguishing the difference of the ray performances and aberrations, especially in the longitudinal direction, between the rotationally and non-rotationally symmetric systems. Then, the calculation of the surface-by-surface generalized chromatic aberrations is carried out in both ray- and wavefront-based models. In both methods, the fields are considered individually so that the surface contributions can be represented as a full-field display. The intrinsic and induced chromatic aberrations as well as the chromatic pupil aberration are discussed. Four examples are given to prove the theory. The example of double Gauss system shows the consistency of the generalized methods with Seidel aberrations in the lowest aberration order. The example of stereomicroscope further double-check the correctness of both ray- and wavefront-based methods. An anamorphic example proves the necessity of considering chromatic pupil aberration in some systems. The last example of a HMD system reveals the chromatic

aberrations of the mirror surfaces by dividing the surface contributions into intrinsic and induced components.

Tolerancing is an essential final step in optical design towards a meaningful manufacturing. However, the tolerancing of freeform surfaces in non-rotationally symmetric systems has not been systematically studied in depth yet. The manufacturing surface errors can be decomposed into amplitudes and phases in the frequency domain. This work characterizes the ACFs of the unwrapped phase patterns in different frequency ranges. By investigating three examples, it has been found that the width of ACF of the statistical errors is significantly thinner than the deterministic errors, which could be used to search for the boundary frequency between the deterministic and statistical surface errors. The concept is applied in both diamond-turned and polished surfaces for demonstration. Then, the workflow of modelling a synthetic freeform surface is put forward by generating and overlaying the LSF, MSF and HSF surface errors. Using the HMD system as an example, the effects of the individual surface errors from LSF to HSF ranges and the total freeform surface error on the spot, PSF and MTF are studied. The small angle scattering of the MSF errors and the wide angle scattering of the HSF errors can be clearly distinguished.

In general, the theories and methods put forward in this thesis can be applied to optical systems and surfaces with arbitrary symmetry. The obscuration elimination algorithm significantly reduces the human involvement and time cost in finding a suitable initial system structure, which could be coupled in sophisticated automatic optimization algorithm in the future. The definition and calculation methods of the generalized chromatic aberrations fill in the lack of a chromatic aberration theory in non-rotationally symmetric optical systems. It gives the optical designer a direct visualization about the distribution of the chromatic aberrations in a general optical system with refractive surfaces, which is helpful in analyzing and compensating the generalized chromatic aberrations. The tolerancing of freeform surface is additionally a powerful tool to predict the influence of manufactured freeform surface errors on real optical systems.

Appendix A

There are in total three steps to transfer ray vectors from one surface to another. First, the ray propagates behind a surface in the free space can be described by the matrix [A-1]

$$\mathbf{M}_{t,i} = \begin{bmatrix} 1 & 0 & \frac{d_{i-1}}{n_{i,\lambda_0}} & 0 \\ 0 & 1 & 0 & \frac{d_{i-1}}{n_{i,\lambda_0}} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.1})$$

where d_i is the chief ray path length of λ_0 between the surface $i-1$ and surface i , n_{i,λ_0} is the refractive index of the central wavelength before the surface i .

Then, the rotation of the coordinate systems between the two surfaces is governed by [A-1]

$$\mathbf{M}_{r,i} = \begin{bmatrix} \cos \xi_i & \sin \xi_i & 0 & 0 \\ -\sin \xi_i & \cos \xi_i & 0 & 0 \\ 0 & 0 & \cos \xi_i & \sin \xi_i \\ 0 & 0 & -\sin \xi_i & \cos \xi_i \end{bmatrix}. \quad (\text{A.2})$$

Last, the transfer matrix at surface is [A-1]

$$\mathbf{M}_{t,i} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{\cos \theta'_i}{\cos \theta_i} & 0 & 0 \\ -N_i^* c_{02,i} \cos \theta_i \cos \theta'_i & -N_i^* c_{11,i} \cos \theta'_i & 1 & 0 \\ -N_i^* c_{11,i} \cos \theta_i & -N_i^* c_{20,i} & 0 & \frac{\cos \theta_i}{\cos \theta'_i} \end{bmatrix} \quad (\text{A.3})$$

where

$$N_i^* = \frac{n'_{i,\lambda_0} \cos \theta'_i - n_{i,\lambda_0} \cos \theta_i}{\cos \theta_i \cos \theta'_i} \quad (\text{A.4})$$

and the expressions of c_{02} , c_{11} , c_{20} can be found in the Eq. (2.11).

Therefore, for each surface i , the transformation matrix \mathbf{M}_i is the multiplication of the above three matrices

$$\mathbf{M}_i = \mathbf{M}_{r,i} \mathbf{M}_{t,i} \mathbf{M}_{t,i}. \quad (\text{A.5})$$

Appendix B

Table B Comparison of the transverse chromatic aberration in the stereomicroscope between the ray- and wavefront-based method, the errors are relative to the ray-based method per surface.

| Surface | X-direction | | | Y-direction | | |
|------------|----------------------|----------------------|-------|----------------------|----------------------|-------|
| | Ray-based | Wavefront-based | Error | Ray-based | Wavefront-based | Error |
| | [D _{Airy}] | [D _{Airy}] | [%] | [D _{Airy}] | [D _{Airy}] | [%] |
| 1 | 1.67 | 1.69 | 1.20 | 1.69 | 1.71 | 1.18 |
| 2 | -0.41 | -0.42 | 2.44 | -0.42 | -0.43 | 2.38 |
| 3 | -1.24 | -1.25 | 0.81 | -1.26 | -1.27 | 0.79 |
| 4 | 1.38 | 1.39 | 0.72 | 1.40 | 1.41 | 0.71 |
| 5 | 1.22 | 1.23 | 0.82 | 1.23 | 1.25 | 1.63 |
| 6 | -2.69 | -2.72 | 1.12 | -2.72 | -2.76 | 1.47 |
| 7 | 5.47 | 5.51 | 0.73 | 5.53 | 5.59 | 1.08 |
| 8 | -6.16 | -6.19 | 0.49 | -6.22 | -6.28 | 0.96 |
| 9 | 0.58 | 0.58 | 0.00 | 0.59 | 0.59 | 0.00 |
| 10 | 2.76 | 2.77 | 0.36 | -5.30 | -5.36 | 1.13 |
| 11 | 0.43 | 0.42 | -2.33 | -4.72 | -4.72 | 0.00 |
| IMA | 3.01 | 3.02 | 0.33 | -10.19 | -10.25 | 0.59 |

References

- [1-1] Mait, J.N., *A history of imaging: revisiting the past to chart the future*, Optics and Photonics News, **17**(2): 22-27, 2006.
- [1-2] Hamilton, W.R., *The Mathematical Papers of Sir William Rowan Hamilton*, Vol. 3: CUP Archive, 1931.
- [1-3] Bruns, H., *Das eikonol*, Vol. 35: S. Hirzel, 1895.
- [1-4] Herzberger, M., *Modern geometrical optics*, Huntington, 1980.
- [1-5] Seidel, L., *Zur Dioptrik. Über die Entwicklung der Glieder 3ter Ordnung welche den Weg eines ausserhalb der Ebene der Axe gelegene Lichtstrahles durch ein System brechender Medien bestimmen*, von Herrn Dr. L. Seidel, Astronomische Nachrichten, **43**: 289, 1856.
- [1-6] Kurkjian, C.R., Prindle W.R., *Perspectives on the history of glass composition*. Journal of the American Ceramic Society, **81**(4): 795-813, 1998.
- [1-7] King, H.C., *The history of the telescope*, Courier Corporation, 2003.
- [1-8] Schwarzschild, K., *Untersuchungen zur geometrischen Optik: Einleitung in die Fehlertheorie optischer Instrumente auf Grund des Eikonalbegriffs. I*, Vol. 1: Weidmann, 1905.
- [1-9] DeHoog, E. A., *Anamorphic zoom lens based on rotating cylindrical lenses*, Optics express **29**(8): 12206-12214, 2021.
- [1-10] Ge, X., Chen, S., Zhang, Y., et al., *Broadband astigmatism-corrected spectrometer design using a toroidal lens and a special filter*, Optics & Laser Technology, **65**: 88-93, 2015.
- [1-11] Zhong, Y., Tang, Z., Gross, H., *Correction of 2D-telecentric scan systems with freeform surfaces*, Optics express, **28**(3): 3041-3056, 2020.
- [1-12] Wei, L., Li, Y., Jing, J., Feng, L., Zhou, J., *Design and fabrication of a compact off-axis see-through head-mounted display using a freeform surface*, Optics express, **26**(7): 8550-8565, 2018.
- [1-13] Alvarez, L.W., *Two-element variable-power spherical lens*, U.S. Patent No. 3,305,294, 1967.
- [1-14] Schwiegerling, J., Paleta-Toxqui, C., *Minimal movement zoom lens*, Applied optics, **48**(10): 1932-1935, 2009.
- [1-15] Xu, C., Cheng, D., Wang, Y., *Automatic obscuration elimination for off-axis mirror systems*, Applied optics, **56**(32): 9014-9022, 2017.

- [1-16] Thompson, K.P., *Aberration fields in tilted and decentered optical systems*, Ph.D. Thesis, The University of Arizona, 1980.
- [1-17] Cao, Y., Lu, L., Deng, Z., *Chromatic aberration of plane-symmetric optical systems*. Applied optics, **58**(2): 227-233, 2019.
- [1-18] Ni, J., Yang, T., Liu, Y., Cheng, D., Wang, Y., *Description and tolerance analysis of freeform surface figure error using specific-probability-distributed Gaussian radial basis functions*, Optics express, **27**(22): 31820-31839, 2019.
- [1-19] Stock, J., Broemel, A., Hartung, J., Ochse, D., Gross, H., *Description and reimplementaion of real freeform surfaces*, Applied optics, **56**(3): 391-396, 2017.
- [1-20] He, L., Evans, C.J., Davies, A., *Two-quadrant area structure function analysis for optical surface characterization*, Optics express, **20**(21): 23275-23280, 2012.
- [1-21] Aryan, H., Boreman, G.D., Suleski, T.J., *Simple methods for estimating the performance and specification of optical components with anisotropic mid-spatial frequency surface errors*, Optics express, **27**(22): 32709-32721, 2019.
- [1-22] Huang, C.-Y., Liang, R., *Modeling of surface topography in single-point diamond turning machine*, Applied optics, **54**(23): 6979-6985, 2015.
- [2-1] Gross, H., *Handbook of Optical Systems*, Vol. 1: Wiley-VCH, 2005.
- [2-2] Forbes, G., *Shape specification for axially symmetric optical surfaces*, Optics express, **15**(8): 5218-5226, 2007.
- [2-3] Abramowitz, M., Stegun, I. A., Romer, R. H., *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Vol. 55: US Government printing office, 1964.
- [2-4] Forbes, G., *Characterizing the shape of freeform optics*, Optics express, **20**(3): 2483-2499, 2012.
- [2-5] Broemel, A., Lippmann, U., Gross, H., *Freeform surface descriptions. Part I: mathematical representations*, Advanced Optical Technologies, **6**(5): 327-336, 2017.
- [2-6] Gross, H., *Handbook of Optical Systems*, Vol. 4: Wiley-VCH, 2005.
- [2-7] Ma, Y., *Design of a 16.5 megapixel camera lens for a mobile phone*, Open Access library journal, **2**(03): 1, 2015.
- [2-8] Gross, H., *Handbook of Optical Systems*, Vol. 3: Wiley-VCH, 2005.
- [2-9] Rolf Wartmann, *Anastigmatic Anamorphic Lens*, US patent No. 6,310,731, 2001.
- [2-10] Zhong, Y., *Optical design methods of non-rotationally symmetric optical systems with freeform surfaces*, Ph.D. Thesis, Friedrich-Schiller-Universität Jena, 2018.
- [2-11] Tang, Z. Gross, H., *Improved correction by freeform surfaces in prism spectrometer concepts*, Applied Optics, **60**(2): 333-341, 2021.

References

- [2-12] Buchroeder, R. Leonard, A., *Wide-field tilted-component telescope: a Leonard extended Yolo all-reflecting system*, Applied Optics, **11**(7): 1649-1651, 1972.
- [2-13] Fallah, H.R., *Astigmatic foci along a skew ray*. Journal of Modern Optics, **51**(13): 1963-1982, 2004.
- [2-14] Buchroeder, R.A., *Tilted component optical systems*, Ph.D. Thesis, The University of Arizona, 1976.
- [2-15] Shack, R.V. Thompson, K.P., *Influence of alignment errors of a telescope system on its aberration field*, Optical Alignment I, SPIE, **251**: 146-153, 1980.
- [2-16] Fuerschbach, K., Rolland, J.P., Thompson, K.P., *Theory of aberration fields for general optical systems with freeform surfaces*, Optics express, **22**(22): 26585-26606, 2014.
- [2-17] Zhu, J., Hou, W., Zhang, X., Jin, G., *Design of a low F-number freeform off-axis three-mirror system with rectangular field-of-view*, Journal of Optics, **17**(1): 015605, 2014.
- [2-18] Rodgers, J.M., *Control of packaging constraints in the optimization of unobscured reflective systems*, Reflective Optics, International Society for Optics and Photonics, **751**: 143-149, 1987.
- [2-19] Nie, Y., Duerr, F., Ottevaere, H., *Automated design of unobscured four-mirror freeform imaging systems*, Freeform Optics, Optical Society of America: FM3B. 5., 2019.
- [2-20] Papa, J.C., Howard, J.M., Rolland, J.P., *Survey of the four-mirror freeform imager solution space*, Optics Express, **29**(25): 41534-41551, 2021.
- [2-21] Hartung, J., Beier, M., Risse, S., *Novel applications based on freeform technologies*, Optical Fabrication, Testing, and Metrology VI, International Society for Optics and Photonics, **0692**: 106920K, 2018.
- [2-22] Pertermann, T., Hartung, J., Beier, M., et al., *Angular resolved power spectral density analysis for improving mirror manufacturing*, Applied Optics, **57**(29): 8692-8698, 2018.
- [2-23] Youngworth, R.N., Gallagher, B.B., Stamper, B.L., *An overview of power spectral density (PSD) calculations*, Optical Manufacturing and Testing VI, SPIE, **5869**: 206-216, 2005.
- [2-24] Parks, R.E., Tuell, M.T., *The structure function as a metric for roughness and figure*, Optical System Alignment, Tolerancing, and Verification X, International Society for Optics and Photonics. **9951**: 99510J, 2016.

- [2-25] Kreis, T., Burke, J., Bergmann, R., *Surface characterization by structure function analysis*, Journal of the European Optical Society-Rapid publications, **9**, 2014.
- [2-26] Gross, H., *Lens Design*, (In preparation).
- [2-27] Gross H., Brömel, A., Beier, M., et al., *Overview on surface representations for freeform surfaces*, Optical Systems Design 2015: Optical Design and Engineering VI, SPIE, **9626**: 195-208, 2015.
- [2-28] Liang, K., *Mathematical Tools for Understanding the Effects of Mid-spatial Frequency Structures on Freeform Surfaces*. Ph.D. Thesis, University of Rochester, 2020.
- [2-29] Charnotskii, M., *Comparison of four techniques for turbulent phase screens simulation*, JOSA A, **37**(5): 738-747, 2020.
- [2-30] Harvey, J.E., *Integrating optical fabrication and metrology into the optical design process*, Applied optics, **54**(9): 2224-2233, 2015.
- [2-31] Tamkin, J.M., Milster, T.D., Dallas, W., *Theory of modulation transfer function artifacts due to mid-spatial-frequency errors and its application to optical tolerancing*, Applied optics, **49**(25): 4825-4835, 2010.
- [2-32] Liang, K., Alonso, M.A., *Understanding the effects of groove structures on the MTF*, Optics Express, **25**(16): 18827-18841, 2017.
- [2-33] Harvey, J.E., *Light-scattering characteristics of optical surfaces*, Stray Light Problems in Optical Systems, International Society for Optics and Photonics, **107**: 41-47, 1977.
- [4-1] Welford, W.T., *Aberrations of optical systems*, Routledge, 2017.
- [4-2] OpticStudio, *Double Gauss 5 degree field*, 2021.
- [4-3] Pretorius, M., *Anamorphic imaging objective*, US Patent No. 8,174,773, 2012.
- [4-4] Chen, B., Herkommer, A.M., *Alternate optical designs for head-mounted displays with a wide field of view*, Applied optics, **56**(4): 901-906, 2017.
- [5-1] Sekman, Y., Trost, M., Larmers, T., Hartung, J., Risse, S., Schröder, S., *Unified representation of mid-spatial frequency errors*, (In preparation), 2022.
- [5-2] Martinez-Carranza, J., Falaggis, K., Kozacki, T., *Fast and accurate phase-unwrapping algorithm based on the transport of intensity equation*, Applied Optics, **56**(25): 7079-7088, 2017.
- [5-3] Cohen, L., *The generalization of the wiener-khinchin theorem*, in Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181), IEEE, **3**: 1577-1580, 1998.

References

- [5-4] Zelen, M., Severo, N.C., *Probability functions*. Handbook of mathematical functions, **5**: 925-995, 1964.
- [5-5] Harvey, J.E., *Parametric analysis of the effect of scattered light upon the modulation transfer function*, Optical Engineering, **52**(7): 073110, 2013.
- [5-6] Personal communication with Fraunhofer Institute for Applied Optics and Precision Engineering IOF, Jena, Germany.
- [A-1] Araki, K., *Analysis of off-axial optical systems (2)*, Optical Review, **7**(4): 326-336, 2000.

List of Figures

| | | |
|-----------|---|----|
| Fig. 2-1 | Geometry of optical surfaces and systems. The surface shape and the bending of the optical axis together determine the geometry of an optical system. | 7 |
| Fig. 2-2 | Comparison between (a) strong and (b) mild aspheres regarding basic shape and orthogonality. | 9 |
| Fig. 2-3 | Zernike Fringe polynomials in 2D sorting (m and n numbering on the edge) and 1D sorting (numbering on the pattern). | 10 |
| Fig. 2-4 | Three examples of double-plane symmetric systems: (a) laser shaping system [2-8], (b) anamorphic photographic lens [2-9], (c) 2D telecentric scan system [1-11]. | 12 |
| Fig. 2-5 | Yolo telescope with two mirrors. The OAR is folded in three-dimensional (3D) space with three segments. Two pairs of the neighboring segments construct two planes in blue and orange to indicate the 3D folding more clearly. | 14 |
| Fig. 2-6 | Separation of tangential and sagittal focal points due to the inclined incidence angle I . L'_T and L'_S denote the tangential and sagittal focal lines. | 15 |
| Fig. 2-7 | An astigmatic ray bundle going through a freeform surface with an infinite small numerical aperture. The surface segment illuminated by the ray bundle is approximated by a toroidal surface. | 16 |
| Fig. 2-8 | Relations between wave aberration, transverse aberration, and longitudinal aberration. | 18 |
| Fig. 2-9 | Longitudinal chromatic aberration $\Delta s'_\lambda$ and transverse chromatic aberration $\Delta y'_\lambda$ between two wavelengths (blue and red). | 19 |
| Fig. 2-10 | Displacement of the effective field center of the surface i in the image plane [1-16]. | 20 |
| Fig. 2-11 | Construction of ray-quadrangles in a three-mirror telescope. | 21 |
| Fig. 2-12 | Two situations of obscuration between the ray-quadrangles a (red) and b (blue): (1) the vertex P is inside a , (2) one side of b cross the hard side of a at the point X | 22 |
| Fig. 2-13 | Manufacturing surface error of a diamond-turned surface. | 23 |
| Fig. 2-14 | 1D integrated PSD in the log-log diagram [2-26, 2-27]. | 25 |
| Fig. 2-15 | PSD of the Zernike fringe polynomial terms with radial orders $n = 4, 8, 12, 16$ and 20 [2-26]. | 26 |
| Fig. 2-16 | 1D profile fitting with equally sampled RBFs [2-26]. | 27 |

List of Figures

| | | |
|-----------|---|----|
| Fig. 2-17 | Random surface errors (a) and its 1D integrated PSD (b) generated by the direct method. Random surface errors (c) and its 1D integrated PSD (d) generated by the convolution method. | 29 |
| Fig. 2-18 | Change of the nominal PSF (dashed curve) due to (a) LSF, (b) MSF, and (c) HSF errors [2-30]. | 30 |
| Fig. 3-1 | 3D Yolo telescope with M1 and M2 rotate freely in 3D space. The propagation of the tangential and sagittal ray bundles are drawn in red and green respectively. As a result, the two neighboring ray-quadrangles (blue and yellow) are no longer in the same plane. | 33 |
| Fig. 3-2 | Process of deriving the surface and ray bundle relation from 3D to 2D. The dashed black curves show the boundary of the mirrors. The numbers denote the sequence of the sides. | 34 |
| Fig. 3-3 | Enumeration of the position relation between a ray-quadrangle and a line. | 35 |
| Fig. 3-4 | The cases that the surface blocks the ray bundle in 3D but there is no truncation between line and quadrant in 2D. The clearance is set to be zero in the drawing. | 36 |
| Fig. 3-5 | Circle b is in the “absolute non-intersection region”. The clearance is set to be zero in the drawing. | 37 |
| Fig. 3-6 | Ways to check the intersection between mirrors in (a) and the intersection between a ray bundle and a mirror in (b). | 38 |
| Fig. 3-7 | Steps to calculate $Erf_{3D}(a, b)$ | 39 |
| Fig. 3-8 | Workflow of the automatic obscuration elimination process. | 40 |
| Fig. 3-9 | (a) Original three-mirror telescope with all surfaces aligned on the common optical axis. (b) The unobscured 3D system structure after optimization. | 41 |
| Fig. 3-10 | Evolution of Erf_{3D} during optimization. | 42 |
| Fig. 3-11 | Original two-mirror telescope with all surfaces aligned on the common optical axis. | 42 |
| Fig. 3-12 | M2 is rotated around y_c -axis. The azimuthal angles between the local ray-caustic coordinate system (x_c, y_c, z_c) and local ray-incidence coordinate system (x, y, z) in both object and image space are 90° | 44 |
| Fig. 3-13 | Erf as a function of M1 tilt around x-axis and M2 tilt around y-axis. The value of Erf between 0 and 100 are mapped with color bar, all values larger than 100 are shown in the same color to represent obscured area. The red lines mark the valleys of zero astigmatism. | 45 |

| | | |
|-----------|---|----|
| Fig. 3-14 | Change of the error function Erf during optimization for 200 iterations. The best result region is marked with red circle. | 46 |
| Fig. 3-15 | (a) Layout of unobscured two-mirror Yolo telescope after optimization. (b) The corresponding spot diagram. | 46 |
| Fig. 3-16 | The RMS spot radius map of the whole FOV before (a) and after (b) local optimization. | 47 |
| Fig. 3-17 | The global tangential (a) and sagittal (b) view of the final Yolo telescope. | 47 |
| Fig. 4-1 | Astigmatic imaging of a single field and a single wavelength. | 49 |
| Fig. 4-2 | Astigmatic imaging of the whole FOV for a single wavelength. | 50 |
| Fig. 4-3 | The generalized chromatic aberrations in a non-rotationally symmetric refractive optical system. | 50 |
| Fig. 4-4 | Propagation of the chromatic intermediate image points of OAR or CR (a) from surface to surface and (b) in the final image space in the sagittal cross section. Blue and red indicate the physical quantities of λ_1 and λ_2 respectively. The purple points show the virtual image points and their transformation to the image space. | 53 |
| Fig. 4-5 | Propagation of ray vectors which are generated at each surface to the final image plane to get the transverse chromatic aberration. $\Delta\vec{H}$ and \vec{T} are in the local ray-incidence coordinate system of the last surface and coordinate system in the image plane, respectively. The image plane is defined with respect to the central wavelength λ_0 | 55 |
| Fig. 4-6 | Construction of reference sphere for (a) tilt and (b) defocus. | 58 |
| Fig. 4-7 | Total chromatic wavefront aberration behind each surface. | 59 |
| Fig. 4-8 | Displacement of pupils between two wavelengths due to the chromatic pupil aberration. | 61 |
| Fig. 4-9 | Separated chief ray paths for two wavelengths and their own physical quantities. | 61 |
| Fig. 4-10 | Layout of the double Gauss system. | 62 |
| Fig. 4-11 | Surface-by-surface comparison of (a) the longitudinal chromatic aberration and (b) the transverse chromatic aberration generated by the three methods. | 63 |
| Fig. 4-12 | Layout of the stereomicroscope for one eye channel. | 64 |
| Fig. 4-13 | Surface-by-surface chromatic image variations calculated by ray- and wavefront-based methods. | 65 |
| Fig. 4-14 | Surface-by-surface transverse chromatic aberrations between the eF'-lines calculated by ray- and wavefront-based methods. | 65 |

List of Figures

- Fig. 4-15 (a) Surface-by-surface chromatic image variations of the on-axis field. The full-field display of surfaces (b) 8, (c) 11 and (d) image plane are shown as the difference between the on-axis field and other fields. The dashed curves mark the fields, which have the same contribution as the on-axis field..... 66
- Fig. 4-16 Full-field display of the transverse chromatic aberration between the (a) eF' -lines and (b) eC' -lines..... 67
- Fig. 4-17 Full-field display of the transverse chromatic aberration between the eF' -lines after removing the decentering effect for (a) the summation of surfaces from 1-9, (b) surface 10, (c) surface 11 and (d) the image plane. 67
- Fig. 4-18 3D layout of the anamorphic photographic lens..... 68
- Fig. 4-19 Full-field display of the chromatic image variation in the image plane. Green circles indicate the nodal points without chromatic image variation. Red means negative, blue means positive values..... 69
- Fig. 4-20 Full-field displays of chromatic image variation on (a) freeform surface 13 and (b) aspherical surface 19 shown as the difference between the on-axis field and other fields. 70
- Fig. 4-21 Surface contributions to the chromatic image variation for the outermost field. 70
- Fig. 4-22 Full-field display of the transverse chromatic aberration in the image plane. 71
- Fig. 4-23 Full-field displays of the transverse chromatic aberration on (a) freeform surface 13 and (b) aspherical surface 19. 71
- Fig. 4-24 Surface contributions to the transverse chromatic aberration between the eF' -lines for the outermost field corner in (a) x - and (b) y -direction..... 72
- Fig. 4-25 Layout of the HMD system. 73
- Fig. 4-26 Surface contributions to the chromatic image variation for the on-axis field. 74
- Fig. 4-27 Full-field display of the chromatic image variation plotted as the difference to the on-axis field on each surface (a)-(d) and in the image plane (e). 75
- Fig. 4-28 Intrinsic, induced and total contributions on each surface for the (a) on-axis field and (b) outermost field. 76
- Fig. 4-29 Full-field display of transverse chromatic aberrations in the image plane..... 76
- Fig. 4-30 Full-field transverse chromatic aberration drawn for each surface. 77
- Fig. 5-1 (a) Surface error described by a Gaussian function in the spatial domain, and its (b) PSD and (c) unwrapped phase in the frequency domain. 79
- Fig. 5-2 (a) 2D PSD with the half width of $4/w$, which is a little bit large than $4w'$ and (b) the corresponding unwrapped phase pattern after removing tilt..... 79
- Fig. 5-3 ACF of the unwrapped phase pattern for seven frequency regions..... 80

| | | |
|-----------|--|----|
| Fig. 5-4 | (a) HWHM and (b) normalized HWHM in the ACF of unwrapped phase pattern. The frequency region in which the phase pattern contains meaningful signals is shadowed in grey..... | 81 |
| Fig. 5-5 | (a) Cosine ripples with a period of 0.5mm in the spatial domain, and their (b) PSD and (c) unwrapped phase in the frequency domain..... | 81 |
| Fig. 5-6 | (a) HWHM and (b) normalized HWHM in the ACF of unwrapped phase pattern. The frequency region in which the phase pattern contains meaningful signals is shadowed in grey..... | 82 |
| Fig. 5-7 | 1D PSD of the cosine ripples. The orange triangle shows the broadening of the spike at 2/mm. | 82 |
| Fig. 5-8 | (a) Statistical errors generated by the direct method in the spatial domain, and their (b) PSD and (c) unwrapped phase in the frequency domain..... | 83 |
| Fig. 5-9 | (a) HWHM and (b) normalized HWHM in the ACF of unwrapped phase pattern. | 83 |
| Fig. 5-10 | Diamond-turned surface errors measured by WLI under four magnifications and their PSDs and unwrapped phases in the frequency domain. | 85 |
| Fig. 5-11 | (a) HWHM and (b) normalized HWHM in the ACF of unwrapped phase pattern for the diamond-turned surface measured at four magnifications..... | 85 |
| Fig. 5-12 | Polished surface errors measured by WLI under four magnifications and their PSDs and unwrapped phases in the frequency domain..... | 86 |
| Fig. 5-13 | (a) HWHM and (b) normalized HWHM in the ACF of unwrapped phase pattern for the polished surface measured at four magnifications..... | 87 |
| Fig. 5-14 | Illustration of the relation between PSD and normalized HWHM for (a) the diamond-turned surface and (b) the polished surface. | 87 |
| Fig. 5-15 | Workflow of the freeform surface modeling and system tolerancing..... | 89 |
| Fig. 5-16 | The first freeform surface in the HMD system is tolerated as an example. | 90 |
| Fig. 5-17 | (a) Figure errors and (b) its 1D integrated PSD. | 91 |
| Fig. 5-18 | Extension of the linearly fitted PSD into (a) MSF and (c) HSF regions. Statistical surface errors in (b) MSF and (d) HSF regions are generated accordingly by the direct method..... | 91 |
| Fig. 5-19 | MSF ripples..... | 92 |
| Fig. 5-20 | Spot diagrams of the fields (0°, 0°), (25°, 0°) in the nominal system and perturbed systems with individual surface errors. The width of each spot diagram is 400μm. | 93 |

List of Figures

| | |
|--|----|
| Fig. 5-21 Slope of (a) original surface and (b) LSF errors. The circles in (b) show the range of ray bundle from individual fields on the surface. The fields are sampled with the normalized coordinates 0, 0.7, 1 on both x - and y -axes..... | 93 |
| Fig. 5-22 PSFs of the fields $(0^\circ, 0^\circ)$, $(25^\circ, 0^\circ)$ in the nominal system and perturbed systems. | 94 |
| Fig. 5-23 Cross sections of PSFs along the x -axis for the central field in the nominal system and perturbed systems. | 94 |
| Fig. 5-24 MTFs of the fields $(0^\circ, 0^\circ)$, $(25^\circ, 0^\circ)$ in the nominal system and perturbed systems. T is the tangential cross section, S is the sagittal cross section. | 95 |
| Fig. 5-25 Total synthetic surface errors with zoom-in inspection. The PSD across the whole frequency range is shown in the lower right side. | 96 |
| Fig. 5-26 Spot diagrams of the fields $(0^\circ, 0^\circ)$, $(25^\circ, 0^\circ)$ in the nominal system and the system with total surface errors. | 96 |
| Fig. 5-27 PSFs of the fields $(0^\circ, 0^\circ)$, $(25^\circ, 0^\circ)$ in the nominal system and the system with total surface errors. | 97 |
| Fig. 5-28 Cross sections of PSFs along the x -axis for the central field in the nominal system and the system with total surface errors. | 97 |
| Fig. 5-29 MTFs of the fields $(0^\circ, 0^\circ)$, $(25^\circ, 0^\circ)$ in the nominal system and the system with total surface errors. | 97 |

List of Tables

| | |
|---|-----|
| Table 2-1 Conic constant and surface shape..... | 8 |
| Table 2-2 Some rotationally symmetric optical systems [2-6, 2-7]..... | 11 |
| Table 2-3 Layouts of some plane symmetric optical systems in the tangential plane [2-6, 2-10, 2-11]. The bended OAR is drawn in each layout..... | 13 |
| Table 2-4 Comparison between Seidel aberrations and wave aberrations represented by Zernike polynomials..... | 19 |
| Table 2-5 Mathematical functions of different types of MSF ripples [2-28]. | 27 |
| Table 2-6 Current status of the research for different system types. Green, yellow, and red indicate that the issue is fully, partly, or not solved. The issue of obscuration is unavoidable in rotational and double-plane symmetric optical systems with reflective surface. | 31 |
| Table 3-1 Initial values, optimization boundaries and final values after optimization of the variables for the three-mirror Yolo telescope. | 42 |
| Table 4-1 Aberrations in the longitudinal direction for different types of optical systems. | 51 |
| Table 4-2 Comparison between traditional and generalized chromatic aberrations..... | 52 |
| Table 4-3 Basic data of the double Gauss system. | 63 |
| Table 4-4 Basic data of the stereomicroscope. | 64 |
| Table 4-5 Basic data of the anamorphic photographic lens. The effective F-number is the average of the tangential and sagittal F-numbers..... | 68 |
| Table 4-6 Comparison between original and modified calculations for the outermost field corner..... | 69 |
| Table 4-7 Basic data of the HMD system..... | 73 |
| Table B Comparison of the transverse chromatic aberration in the stereomicroscope between the ray- and wavefront-based method, the errors are relative to the ray-based method per surface. | 101 |

List of Abbreviations

| | |
|------------|--|
| AR | Augmented reality |
| VR | Virtual reality |
| MR | Marginal ray |
| CR | Chief ray |
| LSF | Low spatial frequency |
| MSF | Mid spatial frequency |
| HSF | High spatial frequency |
| FOV | Field of view |
| PSD | Power spectral density |
| RMS | Root-mean-square |
| OAR | Optical axis ray |
| TMA | Three mirror anastigmat |
| HMD | Head-mounted-display |
| 1D, 2D, 3D | One-, two-, three-dimensional |
| NA | Numerical aperture |
| OPD | Optical path difference |
| NAT | Nodal aberration theory |
| FTS | Fast tool servo |
| STS | Slow tool servo |
| CCP | Computer-controlled polishing |
| MRF | Magnetorheological finishing |
| PV | Peak-valley |
| 2D-DFT | Two-dimensional discrete Fourier transform |

| | |
|-------------------|---------------------------------|
| RBFs | Radial basis functions |
| iFFT | Inverse fast Fourier transform |
| PSF | Point spread function |
| MTF | Modulation transfer function |
| STF | Surface transfer function |
| ACV | Autocovariance function |
| EnP | Entrance pupil |
| ExP | Exit pupil |
| RL | Rayleigh unit |
| D_{Airy} | Airy diameter |
| IMA | Image plane |
| TIE | Transport of intensity equation |
| ACF | Autocorrelation function |
| HWHM | Half width at half maximum |
| WLI | White light interferometry |

List of Symbols

| | |
|----------------------------------|--|
| c | Radius of curvature |
| κ | Conic constant |
| \bar{x}, \bar{y} | Normalized Cartesian coordinates on optical surface |
| \bar{r} | Normalized radius on optical surface |
| c_{bfs} | Radius of curvature of the best fitted sphere |
| ϕ_{pr} | Projection factor |
| (r, φ) | Polar coordinates |
| n, n' | Refractive index in the object and image space |
| l, l' | Object and image distance |
| R | Surface radius |
| f | Focal length |
| t, t' | Object and image distance along the chief ray in the tangential plane |
| s, s' | Object and image distance along the chief ray in the sagittal plane |
| I, I' | Ray-incidence and refraction angle |
| (x_s, y_s, z_s) | Local surface coordinate system |
| $c_{\text{max}}, c_{\text{min}}$ | Maximum and minimum principal curvatures |
| η | Azimuthal angle between the x_s -axis and the direction of the maximum principal curvature |
| (x, y, z) | Local ray-incidence coordinate system in the object space |
| (x', y', z') | Local ray-incidence coordinate system in the image space |
| (x_c, y_c, z_c) | Local ray-caustic coordinate system in the object space |
| (x'_c, y'_c, z'_c) | Local ray-caustic coordinate system in the image space |
| θ, θ' | Azimuthal angles between the ray-incidence and ray-caustic coordinate systems |

| | |
|------------------------|--|
| W | Wavefront aberration |
| λ | Wavelength under consideration |
| $\Delta x', \Delta y'$ | Geometrical aberration in x - and y -directions |
| $\Delta s'$ | Longitudinal aberration |
| x_p, y_p | Cartesian coordinates on the exit pupil |
| R_{ref} | Radius of reference sphere in the exit pupil |
| H | Field height in the image plane |
| (ρ, φ) | Radial height and azimuthal angle on the exit pupil |
| $\Delta s'_\lambda$ | Longitudinal chromatic aberration |
| $\Delta y'_\lambda$ | Transverse chromatic aberration |
| \vec{H} | Field vector in the image plane |
| $\vec{\rho}$ | Pupil vector |
| $\vec{\sigma}_i$ | Field shift vector for surface i |
| \vec{H}_{Ai} | Effective field vector for surface i |
| Erf | Error function |
| P_j | Vertex of a ray-quadrangle |
| A | Extended area of a ray-quadrangle |
| X | Cross point between the sides of two ray-quadrangles |
| N_q | Total number of ray-quadrangles in a system |
| Δz | Manufacturing surface error |
| z_{real} | Sag of manufactured surface |
| z_{ideal} | Sag of designed surface |
| U | Fourier transform of surface error |
| ν_x, ν_y | Frequencies |
| A | Surface area |

List of Symbols

| | |
|--------------------------|---|
| D_x, D_y | Surface width in x - and y -directions |
| $\Delta x, \Delta y$ | Sampling spacing in x - and y -directions in spatial domain |
| $\Delta v_x, \Delta v_y$ | Sampling spacing in x - and y -directions in frequency domain |
| σ_{rms} | RMS surface error |
| $G(x, y)$ | RBFs |
| (x_m, y_m) | Sampling grid of RBFs |
| w | Shape factor of RBFs |
| h | Amplitude of ripple functions |
| T | Period of ripple functions |
| (x_c, y_c) | Center of ripple functions |
| ϕ | 2D phase pattern |
| L_c | Correlation length |
| D_{Exp} | Diameter of exit pupil |
| r_a, r_{a+1}, r_b | Radii of spheres which contain the mirrors |
| ω | Weighting factor |
| N | Number of optical surfaces in a system |
| S | Number of surfaces including dummy surfaces, object and image planes, etc., in a system |
| Z_i | Zernike Fringe coefficients |
| h_c | Width of circle of least confusion |
| NA'_s, NA'_t | Image-sided numerical apertures in the sagittal and tangential planes |
| $\Delta s'_{i(j)}$ | Sagittal focal line difference generated by surface i and propagated to surface j |
| ${}^sO'$ | Real sagittal focal point |
| ${}^sV'$ | Virtual sagittal focal point |

| | |
|--|--|
| d | Distance between two surfaces |
| ξ | Azimuthal angle between the incidence planes of the two neighboring surfaces |
| $\Delta l'$ | Chromatic image variation |
| $\Delta \vec{H}$ | Vectorial transverse chromatic aberration in the local ray-incidence coordinate system of the last surface |
| \vec{T} | Vectorial transverse chromatic aberration in the local coordinate system of the image plane |
| $\hat{s}, (\alpha_i, \beta_i, \gamma_i)$ | Direction cosine of CR |
| \hat{e} | Surface normal |
| U_i, V_i | Numerical apertures in the $x'z'$ - and $y'z'$ - planes |
| $\Delta \vec{r}_i$ | Ray vector behind surface i |
| \mathbf{M}_i | Transformation matrix for surface i |
| (x_i'', y_i'', z_i'') | Ray vector intersection point on the image plane from surface i |
| \mathbf{G} | Rotational matrix between the local ray-incidence coordinate system of the last surface and the local coordinate system of the image plane |
| \bar{P} | Chief ray intersection point on the surface |
| Q | Chief ray intersection point on the real wavefront |
| Q_0 | Chief ray intersection point on the ideal wavefront |
| Q_N | Chief ray intersection point on the surface |
| ΔW_i | Wavefront aberration of surface i |
| ε_1 to ε_5 | Wavefront aberration coefficients until 2nd order |
| $W_{\text{tilt}}, W_{\text{defocus}}$ | Tilt and Defocus wavefront aberration |
| $R(x,y)$ | Ripple function |
| $S(x,y)$ | Function of statistical errors |

Acknowledgement

In retrospect, the motivation that I came to Germany in 2014 was to pursue in-depth study in the field of optical design. At the moment when I write down this sentence, I have finished my research work about freeform optical systems as a Ph.D. student. My dream could not come true without the kind help from the people I have met in the past years.

I would like to express my sincere appreciation to my supervisor, Prof. Herbert Gross, who gave me the chance to join his research group and guided me through my master's until the end of my Ph.D. study. I still remember the day we met for the first time in his office when we discussed about the prospect of optical design, the combination of research direction and personal development plan. His words have been kept in my mind. During the past seven years, he always offered me professional academic advice and guidance, took care of my interest and small ideals, and encouraged me when I felt lost. He is such an enthusiastic person toward scientific research and is willing to help others. No matter how busy he was, he kept his office door open for discussion, and even replied to my E-mails and corrected my manuscripts during holidays. I always tried to learn from his personality and his attitude toward work and life, which shaped me to be the person who I am.

Also, I would like to thank my colleagues in the research group. They are all very kind. I benefited a lot from the discussions and from the critical questions they put forward, which helped me formulate my research work more comprehensively. My special thanks go to Dr. Yi Zhong and Dr. Yueqian Zhang, who gave me not only suggestions in research direction but also in life, to Dr. Norman Girma Worku for supporting me a lot in programming, and to Dr. Johannes Hartung and Tom Lammers for answering many questions concerning tolerancing.

Finally yet importantly, I would like to thank my parents, Ping Cai and Xiaorong Li, my partner, Yang Wang, and my friends, especially Yang Xia. They shared my happiness and tough moments in the adventure of my Ph.D. journey. I am deeply grateful for their endless support and encouragement.

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

Cai, D., Gross, H., *Obscuration elimination in three-dimensional non-symmetrical optical systems*, Journal of Physics: Photonics, **1**(4): 044002, 2019.

Cai, D., Gross, H., *Generalized chromatic aberrations in non-rotationally symmetric optical systems–Part I: mathematical approach*, Applied Optics, **60**(21): 6313-6321, 2021.

Cai, D., Gross, H., *Generalized chromatic aberrations in non-rotationally symmetric optical systems–Part II: applications*, Applied Optics, **60**(21): 6322-6330, 2021.

Conference Proceedings

Cai, D., Liu, C., Gross, H., et al., *Higher order aberrations of Alvarez lenses*, SAOT International Conference on Advanced Optical Technologies in Erlangen, 2019.

Talks

Cai, D., Liu, C., Gross, H., et al., *Higher order aberrations of Alvarez lenses*, SAOT International Conference on Advanced Optical Technologies in Erlangen, 2019.