MID-SPATIAL FREQUENCY CHARACTERIZATION AND SPECIFICATION FOR FREEFORM SURFACES USING ZERNIKE POLYNOMIALS

by

Zahra Hosseinimakarem

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Optical Science and Engineering

Charlotte

2017

Approved by:

_____
Dr. Angela Davies

_____
Dr. Christopher Evans

_____
Dr. Thomas Suleski

_____
Dr. Gregory J. Gbur

ABSTRACT

ZAHRA HOSSEINIMAKAREM. Mid-spatial frequency characterization and specification for freeform surfaces.
(Under the direction of DR. ANGELA D. DAVIES and DR. CHRISTOPHER J. EVANS)

Mid-spatial frequency on freeform optical elements induces small-angle scatter and affects the optical performance. Fabrication techniques involved in making freeform surfaces leave tooling marks on the surface due to the sub-aperture nature of the fabrication process. In recent years, there has been a growing need for specification and characterization of the mid-spatial frequencies for freeform surfaces. There is a range of methods to consider for representing the mid-spatial frequency content: the power spectral density (PSD), the structure function (SF), and a polynomial basis representation such as Zernike and Forbes polynomials. Using research discussed here and published discussions from the literature, we compare a Zernike representation to a PSD representation, and structure-function representation in terms of ability to connect to optical performance, ability to connect to the fabrication process, and other practical considerations. Our focus is on investigating a Zernike polynomial representation for quantifying the mid-spatial frequency content in the height maps. We will illustrate how this polynomial representation captures certain characteristics of the mid-spatial frequency texture. A filtering aspect of these polynomials is also explored and optical performance is predicted based on very large orders of Zernike polynomials.

DEDICATION

I dedicate this work and my research during the last six years, to my father and my mother who gave me unconditional love and support from the other side of our planet. They have been a tremendous source of encouragement and stood by my side during difficult times and cheered with me for my accomplishments.

ACKNOWLEDGMENTS

I would like to thank my advisors Dr. Angela Davies and Dr. Christopher Evans. During this research, they have provided excellent guidance, patience, and support. I benefited from their experience, knowledge and insight way more that I could have imagined. Aside from their technical knowledge, their leadership, project management style and humane way of interacting with graduate students have inspired me. I am very grateful!

I am also thankful for the time and support of my committee members. The opportunity to audit the lens design class taught by Dr. Suleski and the discussions during and after the class have been very helpful to understanding the impact of mid-spatial frequency. Dr. Gbur has always been available in providing insight and useful references that helped me understand the underlying hypothesis and theory of some of the mathematical methods related to the content in chapters three and five of this dissertation.

I also would like to thank the Center for Freefrom Optics CeFO for giving me a platform to present my work and receive valuable feedback from knowledgeable people in the field including Dr. Kevin Whiteaker, Dr. Jannick Rolland, and Dr. Julius Muschaweck. This work has benefited from personal communication and many published works of Dr. Greg Forbes. I also like to thank Dr. Joseph Owen and Dr. John Troutman for very helpful discussion about some manufactured parts that I analyzed for this work. Most of the samples used in this work were manufactured in-house or provided by the industrial affiliate members of CeFO. Additionally, I would like to thank Reza Aryan for collaborating with me and useful conversions during this research. At the end, I would like to thank friends and special people who helped me grow during these years and awakened love and passion in me that made completion of this work possible.

This work was partially supported by the NSF I/UCRC Center for Freeform Optics (IIP-1338877 and IIP-1338898)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACF	auto correlation function
DFT	discrete Fourier transform
FFT	fast Fourier transform
ITF	instrument transfer function
MRF	magnetorheological finishing
MSF	mid-spatial frequency
MTF	modulation transfer function
OTF	optical transfer function
PSD	power spectral density
PSF	point spread function
PV	peak to valley
RMS	root mean square
RSS	root sum square
SF	structure function

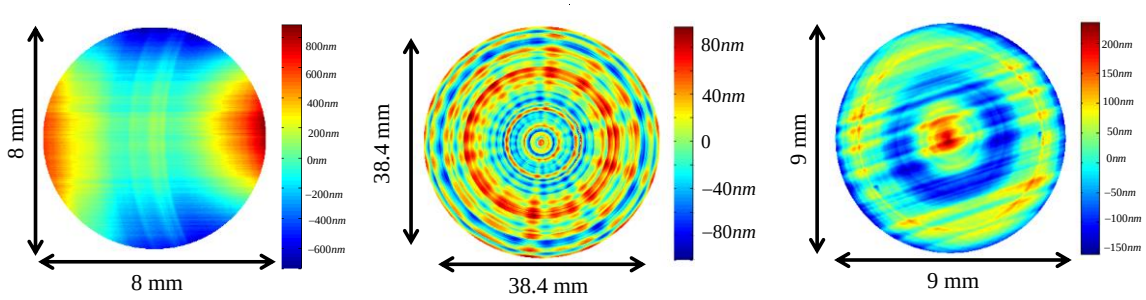
CHAPTER 1 : INTRODUCTION AND LITERATURE REVIEW

1.1 Motivation

Freeform and aspheric optics provide new degrees of freedom for correcting aberrations and open the door to the next generation of innovation in the fields of imaging, illumination, visualization and sensing [1-4]. Freeform optics can have an arbitrary shape and can be described as surfaces with no axis of rotational symmetry [5]. In optical systems, they enable designers to obtain better performance such as simultaneously achieving smaller spot size, increased depth of field and smaller aberrations with fewer optical components in a smaller package [6]. For example, all reflective, off axis, unobscured optics are a class of freeform optics and they are key to the future of EUV lithography [7].

However, the fabrication techniques involved can leave significant millimeter-scale tooling marks, or mid-spatial frequency errors, on the surface because of the sub-aperture nature of the fabrication process. This texture may be difficult to remove especially on hard ceramic materials such as Silicon Carbide (SiC) [8]. Signatures of these structured errors are tool-specific. For example, single point diamond turning leaves rotationally symmetric ring-like structures [9] and raster milling tends to leave linear texture. The geometry of the grinding or polishing machine is not the only cause for generation of mid-spatial frequencies. Overlap in the programmed path of the polishing tool, motion control instability, shape of the motion path, direction of the feed, vibration, tool wear instability and workpiece deformation are other causes of mid-spatial frequency errors [8].

One can view this texture as a disordered diffraction grating on the surface of the optic. Figure 1.1 shows examples of mid-spatial frequency (MSF) signatures for a range of fabrication processes. Figure 1.1 (a) illustrates the surface map of a germanium sphere that was fabricated using precision raster milling. Aside from the low order astigmatism, this surface has some arcs that are due to the change of direction of the tool tip and raster patterns from the overlap in the programmed tool path. Figure 1.1 (b) shows an aluminum diamond turned part where the spokes patterns are associated with the poles in the motor. Figure 1.1 (c) is a chalcogenide glass IRG 26 sphere that was diamond milled and has some raster-like texture and some larger rings due to the thermal cycles of the tool.



(a) (b) (c)

Figure 1.1 Mid-spatial frequency representations on different surfaces. (a) Diamond milled germanium part. (b) Diamond turned aluminum part. (c) Diamond milled infrared chalcogenide glass (IRG 26).

Mid-spatial frequency error leads to small-angle scatter in the transmitted or reflected optical beam [10, 11], and is a concern in many applications. For example, in high power solid state laser systems, initial perturbations evolve into filaments under the mechanism of small-scale self-focusing and finally can damage the optical element [12]. In NIF beamlets, the mid-spatial frequencies on the KDP crystal of the Pockels cell cause small-angle scatter which in high laser intensities may seed beam break up and lead to damage of optical elements due to intensity-dependent part of the refractive index [13]. High-

performance imaging systems are optical systems with small RMS spot size, low aberrations and large Strehl ratio [14]. In high-performance optical systems, MSF error affects the quality of the imaging system and degrades the image contrast, causing image artifacts such as ghosts and local flare [15, 16]. In lithography tools, flare is a significant problem associated with MSF [17]. Flare is wavelength dependent and scales as $1/\lambda^2$. For the same MSF structure, the flare is 200 times larger for EUV systems (13.5 nm wavelength) compared to 193 nm systems [18]. Therefore, it is important to have a clear way of specifying and quantifying MSF texture on optical surfaces to facilitate the specification and acceptance testing of these components.

1.2 What is mid-spatial frequency

The departure of a surface from the design shape can be decomposed into three different spatial frequency regions, with different terminology in optical and mechanical applications. The low-frequency content of the surface is known as form (optical term) or figure (mechanical term) that can be described by low order aberration terms and consist of low-frequency errors (few cycles across aperture). The high-frequency content is known as finish or roughness. Roughness is typically associated with a process signature (for example from grinding or lapping) or uncontrolled sources such as transient seismic vibration. The region between figure and roughness is known as mid-spatial frequency (sometimes called ripple or waviness). These are typically periodic errors that reflect a characteristic of the fabrication technique. For example, in diamond turning machines with a leadscrew, straightness errors at the leadscrew pitch appear in the manufactured surface. In MRF surfaces (Magnetorheological finishing), there is typically a residual signature with a surface wavelength directly related to the size of the removal footprint [19].

The cut-off frequency for each spatial frequency region is defined by the application. For optical elements, this cut-off is defined by the incident wavelength, the size of the aperture and the location of the surface from the stop [11].

A laser Fizeau interferometer can measure spatial periods as small as 100 μm depending on the zoom, the objective lens, and size of the sample [17]. The measuring capability of an interferometer is limited by its optical transfer function (OTF) as well as the pixel density of its CCD. The instrument transfer function (ITF) describes the system's ability to reproduce measurements in terms of spatial frequency. An excellent explanation and calculation of ITF for surface profiling interferometers can be found in [20]. ITF is commonly used for linear systems and their frequency analysis. In linear systems, we can propagate two different frequency components and add the results independently. The measurement of height using optical interferometers is a non-linear process. In the limit of small surface variations ($h \ll \lambda/4$), the interferometer's response can be approximated to be linear and is the same as the imaging OTF (Optical Transfer Function). Imaging a sharp reflectivity step is one of the most convenient ways of measuring OTF [20]. The ratio between the Fourier transform of the image and that of the object is the OTF ($ITF(f_{x,y}) = \sqrt{\frac{PSD(f_{x,y})_{measured}}{PSD(f_{x,y})_{idal}}}$) [21]. Figure 1.2 shows the theoretical and experimental ITF curves for a range of microscope objectives. This figure shows by using different objectives, we can span a wide range of frequencies [20]. The black horizontal dashed line shows 70% of the maximum ITF, and indicates the largest frequency accepted by each objective. As the frequencies approach the Nyquist frequency, the ITF decreases. The frequency at which the

ITF drops to ~ 0.1 of its maximum is the limiting resolution of the system, and a higher magnification objective is needed to resolve larger frequencies [22].

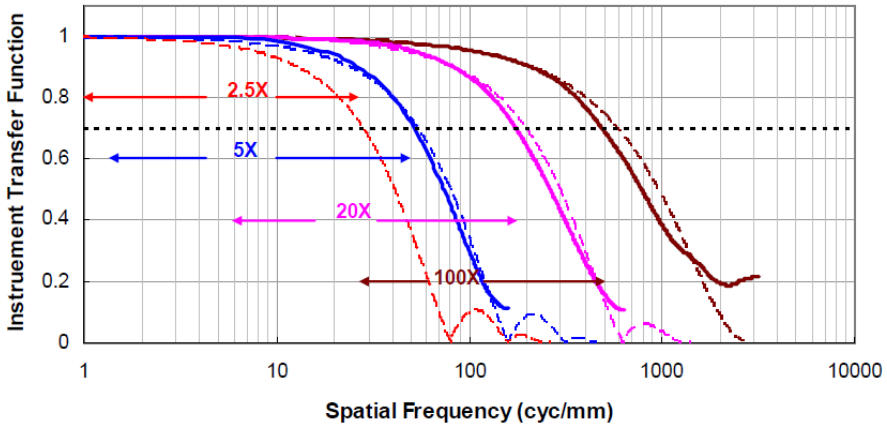


Figure 1.2 The experimental (solid lines) and theoretical (dashed lines) ITF for four different objectives in the interferometric microscope [22].

In the measurement in Figure 1.1 (c), the aperture of the test part is 9 mm and the number of points in the data is 939×939 pixels. That provides a Nyquist limit of $f_{Nyquist} = \frac{N}{2D} = 52.16 \, {}^{1}\!/_{mm}$, corresponding to a spatial period of $0.019mm$. Spatial frequencies above Nyquist may be reported with incorrect amplitude in the absence of appropriate anti-aliasing filters.

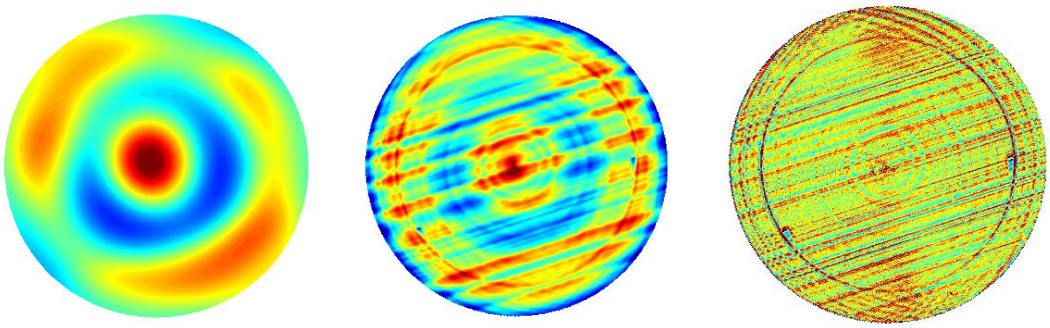
There are different definitions of mid-spatial frequency in the literature. One definition is by the Fresnel length and distance of the surface to the image plane. The Fresnel length of a given spatial frequency is $L_f = \frac{a^2}{\lambda}$ where a is the spatial length and λ is the application wavelength. The Fresnel length (Fresnel distance) provides a coarse criterion for separating near field and far field and the limit where geometrical optics is valid. For $1 \ll \frac{a^2}{\lambda L_f}$ or $L_f \ll \frac{a^2}{\lambda}$ geometrical optics approximations are valid. Beyond the Fresnel

distance, the divergence of the beam becomes significant and we are in the Fraunhofer region.. D. Aikens *et al.* select the low-frequency limit of the mid-spatial frequency errors to be about 5 cycles across the aperture since these errors are associated with low order aberration terms and aberration theory can apply [23]. Low order aberrations such as coma or astigmatism do not affect local image quality metrics but affect the overall image quality metrics such as Strehl ratio. Some authors argue that the selection of the high spatial frequency limit for MSF is arbitrary and based on experience [23, 24]. D. Aikens *et al*. select the high-frequency cut-off of the mid-spatial frequency in an imaging application to be where the Fresnel lengths of the ripples are less than 1/10 of the optical path distance from a given surface to the image plane [23]. Therefore, the high-frequency cut-off can be described as

$$a = \sqrt{\frac{\lambda \times OPD}{10}} \qquad (1.1)$$

where OPD is the optical path length to the image plane. J. Tamkin selects a cut-off frequency of 1/20 of Nyquist to separate mid-spatial frequencies from roughness [24]. He states that this limit places a somewhat arbitrary but useful boundary between MSF errors and roughness [24]. This is the same as assuming the point spread function (PSF) has an isoplanatic patch diameter of 40 Airy disks. Isoplanatic patches are regions where the imaging system is space-invariant and the point source image of the object only changes in location and not in the functional form [25]. In many cases for visible wavelengths, the mid-spatial frequency region is between a few millimeters and 0.1 millimeters [23, 24]. To provide a visual example of different spatial frequency regions of the form (figure), mid-spatial frequency and roughness errors, we have decomposed the SiC sphere in 1.1 (c) into three different spatial frequency bands using Fourier-based filtering capability of the

commercial software MetroProX (MX v6.3.0.4). The results are shown in Figure 1.3.



(a) (b) (c)

Figure 1.3 Surface error of IRG26 decomposed in three different spatial frequency regions. (a) Low spatial frequency region (spatial periods $\Lambda > 3mm$) with RMS= 27.65 nm. (b) Mid-spatial frequency region (spatial periods $0.1mm < \Lambda < 3mm$) with RMS=38.74 nm. (c) High-spatial frequency region (spatial periods $\Lambda < 0.1mm$) with RMS=1.87 nm.

The most common method to specify the surface form and wavefront error is in terms of peak to valley (PV). PV is easy to evaluate but can be highly influenced by outliers [26]. Root mean square error (RMS) is another surface metric that corresponds better with optical performance. To calculate the RMS, each pixel's deviation from the best fit surface is squared, all values are summed, then the square root of the sum is divided by the number of pixels [26]. PVr (Peak to Valley robust) is another metric for optical surface characterization introduced by C. Evans [27]. PVr is defined by $PV_r = PV_{36\ \text{Zernike}} + 3\times\sigma_{36\ \text{Zernike Resid}}$ that is the PV of the surface obtained by least-squares fitting to a 36-term Zernike polynomial and added to three times the RMS of the residual after fitting and removing the 36 terms [26, 27]. In rotationally symmetric systems, the wavefront is described by expanding the wave aberration function in a power series of field and aperture coordinates H, ρ, θ through

$$\begin{aligned} W(H,\rho,\theta) = {} & W_{020}\rho^2 + W_{111}H\rho\cos\theta + W_{040}\rho^4 \\ & + W_{131}H\rho^3\cos\theta + W_{222}H^2\rho^2\cos^2\theta \\ & + W_{220}H^2\rho^2 + W_{311}H^3\rho\cos\theta + O(6) \end{aligned} \qquad (1.2)$$

where ρ is the normalized radial aperture coordinate and H is the normalized image height. $W = W_{020}\rho^2 + W_{111}H\rho\cos\theta$ are first-order aberration terms and $W_{040}\rho^4 + W_{131}H\rho^3\cos\theta + W_{222}H^2\rho^2\cos^2\theta + W_{220}H^2\rho^2 + W_{311}H^3\rho\cos\theta$ are third order aberration terms [28]. Their effect on optical performance is assessed by sequential ray tracing and aberration theory and well described in the literature [14, 29]. First and third-order wavefront properties can be obtained from the Zernike polynomials coefficients from the first nine Zernike terms in the Fringe set (see Table 2.1) [29]. Zernike polynomials will be described in detail in chapter three.

The high spatial frequency components will cause scattering of light and their effect can be simulated based on scattering theory by using non-sequential ray tracing analysis [23]. For smooth surfaces $(\lambda >> 4\pi\sigma\cos\theta_i)$, the reflectance and RMS surface roughness are related through Eqn. 1.3 [30].

$$R_s = R_0\left(\frac{-(4\pi\sigma\cos(\theta_i))^2}{\lambda^2}\right), \qquad (1.3)$$

where R_0 is the reflectance of the smooth surface, R_s is the specular reflectance, σ is the effective Root Mean Square (RMS) surface roughness, θ_i is the angle of incidence relative to the surface normal, and λ is the radiation wavelength. Davies assumes if the surface is smooth, the surface height distribution is Gaussian and most of the light is restricted to the $+/-1^{st}$ diffraction orders $\theta_s \cong \theta_i$ [30]. He also assumes there are no structured patterns in the PSD [30-36]. In summary, if the surface is optically smooth,

homogenous, isotropic and reflective, there is a relationship between RMS of the surface and area PSD [36].

These assumptions are not true for surfaces produced using deterministic polishing techniques with mid spatial frequencies that are usually structured and have a specific direction. Therefore, for MSF texture, RMS is not a sufficient metric for specification. Conventionally polished surfaces have different bands of frequencies. ISO 10110-part 8 (surface texture, roughness, and waviness) provides a mechanism for PSD-based specification. The National Ignition Facility (NIF) used ISO 10110 and PSD analysis on spatially filtered data to specify the wavefront quality of NIF optics in the MSF region [37-39]. This is to ensure that the total scatter loss due to MSF texture does not exceed a certain value [39].

1.3 Overview of mathematical representation options for mid-spatial frequency

Development of freeform optical systems like any other optical system is divided into the major stages of design, fabrication, test, and assembly. The first stage (design) and the last stage (assembly) require a full system analysis while the intermediate 'fabrication' and 'test' stages have a component-level focus. In this section, we introduce methods for quantifying mid-spatial frequencies as a single component representation.

Common methods for representing MSF are power spectral density (PSD), polynomial representation (Zernike and Forbes), structure function, autocorrelation function, wavelet analysis, and slope analysis techniques. The need for MSF characterization and specification is becoming relevant beyond the somewhat narrow field of freeform optics. For example, an array of aspheric elements for computational imaging application poses many of the same manufacturing problems as a freeform surface, hence a need for clear

ways of characterization and specification of their MSF texture. In the next chapter, we will discuss aspects in the context of freeform optics and evaluate/compare each of these techniques based on our evaluation criteria.

1.3.1 PSD

Power spectral density (PSD) is the square magnitude of the Fourier transform of the profile over a certain spatial frequency range. It contains information about the lateral spatial wavelengths on the surface as well as their magnitude since it describes the amount of error in a surface at a particular spatial frequency. PSD is the most common method for characterizing the mid-spatial frequency content from a measurement of a surface [40]. One of the useful parameters that can be extracted from the PSD is the RMS roughness or the RMS over a selected bandwidth. The area under the linear PSD curve within a certain frequency range is the RMS of that frequency range, as shown in Eqn. 1.4 [41].

$$RMS = \sqrt{\int_{\nu_1}^{\nu_2} PSD(\nu)d\nu}. \qquad (1.4)$$

PSD is also useful to distinguish parts with the same RMS. For example, a periodic surface and a random surface may have the same RMS but their PSD will be different. PSD of the periodic surface will have a peak at a certain frequency whereas the PSD of the random surface will have no peaks. Low spatial frequencies must first be removed from the data prior to application of the PSD and there is no agreed upon standard for this high-pass filtering step. A rule of thumb is to remove everything at less than five cycles across the aperture [42]. De-trending of the profile is to remove low-order form errors prior to calculation of PSD to avoid the high-frequency artifacts in the PSD caused by discontinuities at the boundaries. The ASME B46.1-2009 standard has expressions to calculate the PSD. However, the de-trending step is not addressed [43]. Next, the height

data is analyzed with the discrete Fourier transform (DFT) which processes an image as if it were part of a periodic vector (or part of a periodic matrix for area PSD) of identical images. PSD of a discrete profile is calculated through Discrete Fourier Transform (DFT) as

$$PSD(f_m) = \frac{2\Delta x}{N}\left|DFT(f_m)\right|^2, \qquad (0 < m < \frac{N}{2}) \tag{1.5}$$

where f_m is the spatial frequency described as $f_m = \frac{m}{N\Delta x}$, Δx, is the sampling interval of the profile measurements and N is the total number of sampling points. When $m = N/2$ the sampling frequency is $f_m = \frac{1}{2\Delta x}$ which is the Nyquist frequency of the sampled measurement [44]. Most surface profiling instruments today such as Phase Shifting Interferometer (PSI) and Wite Light Scanning Interferometer (SWALI) can capture 2D arrays of surface data. The PSD analysis of linear profiles can be extended to area PSD to perform a better characterization of the surface. The area PSD is the 2D version of Eqn. 1.5 and is defined as

$$\begin{aligned} PSD(f_x, f_y) &= \frac{\Delta x \Delta y}{N_x N_y}\left|DFT(f_x, f_y)\right|^2 \\ &= \frac{\Delta x \Delta y}{N_x N_y}\left|\sum_{n_x=0}^{N_x-1}\sum_{n_y=0}^{N_y-1} e^{-i2\pi\left(f_x n_x \Delta x + f_y n_y \Delta y\right)} z(n_x\Delta x, n_y\Delta y)\right|^2, \end{aligned} \tag{1.6}$$

where $f_x = \frac{m}{N_x\Delta x}$ and $f_y = \frac{n}{N_y\Delta y}$ are the sampling frequency in x and y, Δx and Δy are the sampling intervals in x and y directions and $z(n_x\Delta x, n_y\Delta y)$ is the surface height data [36].

1.3.2 Zernike polynomials

Many optical systems in use today employ imaging elements and pupils that have circular apertures. Therefore, Zernike polynomials are commonly used to mathematically describe optical wavefronts propagating through such systems or components. Zernike polynomials have radial and azimuthal variables that are orthogonal and continuous over the interior of a unit circle [29]. It should be noted that the Zernike polynomials will not be orthogonal over a discrete set of points within a unit circle [45, 46], but are a good approximation for low order terms at reasonable sampling. For example, 21x21 uniformly spaced points provide reasonable sampling for fitting to 36 terms of Zernike polynomials. For more details, please see section 3.5. Zernike polynomials offer distinct advantages over other polynomial basis sets for circular apertures. Using the orthonormal Zernike expansion to describe aberrations offers the advantage that the coefficient or value of each mode is the Root Mean Square (RMS) wavefront error attributable to that mode for a specific representation. Ideally, Zernike coefficients used to mathematically describe a wavefront are independent of the number of polynomials used in the sequence. This condition of independence or orthogonality means that any number of additional terms can be added without changing the coefficients already computed. Coefficients of larger magnitude indicate the greater contribution of that particular mode to the total RMS wavefront error of the system. In chapter three, we will show the results of Zernike analysis for mid-spatial frequency representation. There are many different indexing schemes for Zernike polynomials; the most common are known as the Fringe set (Table 2.1) [47] and the standard set (Table 3.1) [48]. Note that commercial implementations use different sequences of Zernike terms, requiring users to beware of such phrases as “the first 36

Zernikes".

1.3.3 Forbes polynomials

A new set of orthogonal polynomials has been introduced by G. Forbes for characterizing the shape of rotationally invariant aspheres and freeform surfaces [49, 50]. These polynomials are orthogonal in gradient. Describing a freeform surface using Forbes polynomials will require fewer terms and fewer decimal digits compared to standard aspheres [49, 51]. Forbes has shown fitting to extreme orders of these polynomials so that there is one fitted coefficient for every sample point evaluated within the circular aperture and has captured the mid-spatial frequency texture [49].

The sag of a freeform surface in cylindrical polar coordinates as $z = f(\rho, \theta)$ can be expressed in terms of Forbes Q-polynomials as shown in Eqn. 1.7 [49].

$$f(\rho,\theta) = \frac{c\rho^2}{1+\sqrt{1-c^2\rho^2}} + \frac{1}{\sqrt{1-c^2\rho^2}}\{u^2(1-u^2)\sum_{n_f=0} a_n^0 Q_{n_f}^0(u^2) \sum_{m=1} u^m \sum_{n=0} [a_{n_f}^m \cos(m\theta) + b_{n_f}^m \sin(m\theta)] Q_{n_f}^m(u^2)\} \tag{1.7}$$

where $u = \rho/\rho_{max}$, ρ_{max} is the semi-diameter of the cylinder that encloses the surface, c is the curvature of the best-fit sphere, and $Q_{n_f}^m(x)$ is a polynomial of order n_f and the azimuthal order is m. The component within the braces in Eqn. 1.7 shows the departure from the best fit sphere. Analysis of a surface using Forbes Q-polynomials gives access to RMS gradient. Forbes has developed another set of slope orthogonal polynomials that are known as Q^{con} to characterize the shape of rotationally symmetric aspheres based on sag deviation from a close-fitting conic [50]. Q^{con} descriptions are suitable for strong

aspheres. Q^{bfs} (bfs=best-fit-sphere) is a special case of $Q_{n_f}^m$ in Eqn. 1.7 where $m=0$ and is suitable for characterizing mild aspheres [51]. In this dissertation, we refer to the slope-orthogonal $Q_{n_f}^m$ polynomials that are developed to characterize freeform surfaces (Eqn. 1.7).

1.3.4 Structure function

Structure function (SF) is the average height difference squared as a function of separation, which can represent the spatial content of precision surfaces and is mathematically described by Eqn. 1.8.

$$SF = \left\langle \left(z(x,y) - z(x+\tau_x, y+\tau_y) \right)^2 \right\rangle_{x,y} \tag{1.8}$$

where $\langle\ \rangle_{x,y}$ is the spatial average or expectation value over x and y and $z(x+\tau_x, y+\tau_y)$ is the surface error height at a distance (τ_x, τ_y) from position (x, y). The linear SF has been used in astronomy to calculate atmospheric turbulence [52, 53] and specification of large astronomical optics [54, 55]. Linear structure function captures data of all spatial scales but it loses the information on anisotropy. Thomas *et al.* introduced the area structure function [56, 57]. However, they did not address the issue of non-stationary surfaces. The recently introduced two-quadrant area SF by L. He *et al.* characterizes surfaces of arbitrary aperture over any chosen dynamic range while retaining anisotropic information [58-60]. The downside of structure function is that the number of points contributing to the average decreases as τ (separation) reaches towards the aperture size. Also, area SF is not an orthogonal representation. This is discussed in more detail in section 2.2.

1.3.5 Autocorrelation function

The autocorrelation function (ACF) is another mathematical technique to characterize the spatial content of a surface that first has been used in communication theory [61]. The autocorrelation function is the autocovariance function of the surface, normalized by the surface variance. For calculating the linear ACF, the original profile is duplicated and shifted along the original profile with a separation τ, and then the product of the two profiles in the overlapped region is obtained. This calculation can be done for every separation τ. As the duplicated profile is shifted, the overlap area decreases. There are two types of auto-correlation functions, biased ACF and unbiased ACF. The unbiased ACV function is defined as

$$ACV(\tau) = \frac{1}{n-\tau}\sum_{j=1}^{n-\tau} z(j)z(j+\tau), \quad 0 \le \tau \le n-1 \tag{1.9}$$

and the biased ACV is defined by Eqn. 1.10 [62, 63].

$$ACV(\tau) = \frac{1}{n}\sum_{i=1}^{n-\tau} z(i)z(i+\tau). \quad 0 \le \tau \le n-1 \tag{1.10}$$

The biased ACF of a sinusoidal profile shows a decay modulated with the oscillations whereas the unbiased ACF does not show an overall decay. Area ACF contains more information about the surface such as anisotropy and surface texture direction. For stationary surfaces, the area SF is related to ACF through Eqn. 1.11 [64].

$$SF(\tau_x, \tau_y) = 2\sigma^2 \left\{1 - ACF(\tau_x, \tau_y)\right\}. \tag{1.11}$$

1.3.6 Wavelet analysis

Wavelet analysis provides a multiscale representation of the surface textures. Wavelets are band-limited, *i.e.* they are composed of a limited range of several frequencies. They are

localized in both frequency and space (or time). They have variable amplitudes in different spatial (or time) periods.

Wavelet analysis provides a filtering scheme based on localized harmonic analysis using short-term signals called mother wavelets. Wavelet filters can be used to decompose a surface into different spatial scales. This makes it possible to analyze the surface at a particular spatial scale. B. Josso *et al.* have used frequency normalized wavelet transform (FNWT) and applied Daubechies wavelets of order 20 to decompose a surface texture profile in form, waviness, and roughness (Figure 1.4) [65]. The Daubechies wavelets are a class of orthogonal wavelets discovered by Daubechies [66]. They are characterized by a maximal number of vanishing moments for a given support. For each wavelet type of this class, there is a scaling function which generates an orthogonal multiresolution analysis. They have slightly longer supports compared to other wavelets and they sample more values from the signal (or surface depending on application) [67].

Unlike Fourier transform that is a global transform, wavelets are a local transform method [66, 68, 69]. The localization property of wavelet functions makes them a better tool for characterizing local spatial structure compared to Fourier methods such as PSD [70]. To describe a local structure or local defect, we need a large number of polynomials which will make the ray tracing computationally intensive. Therefore, wavelets are advantageous over polynomials for representation of local structures [70].

Wavelets are not suitable for analysis of figure on freeform surfaces or high slope errors due to the assumption that the underlying geometry of the surface is Euclidean or planar [71]. H. S. Abdul-Rahman *et al.* have applied a new generation of wavelets for filtering freeform surface that can filter any arbitrary shape described by a triangular mesh [72].

These new wavelets are produced through a lifting scheme. Lifting scheme starts with a simple wavelet to split the data, then this simple wavelet is lifted and new wavelets are predicted and updated [73].

One dimensional wavelet filters are defined in ISO 16610-29 and areal wavelet filters are defined in ISO 16610-69 for surfaces. So far, only spline wavelets are defined in the standard. In summary, wavelets are preferred over PSD and polynomial methods for characterizing localized mid-spatial frequency errors but the first generation of wavelets are not ideal for characterizing the shape of high slope freeform surfaces.

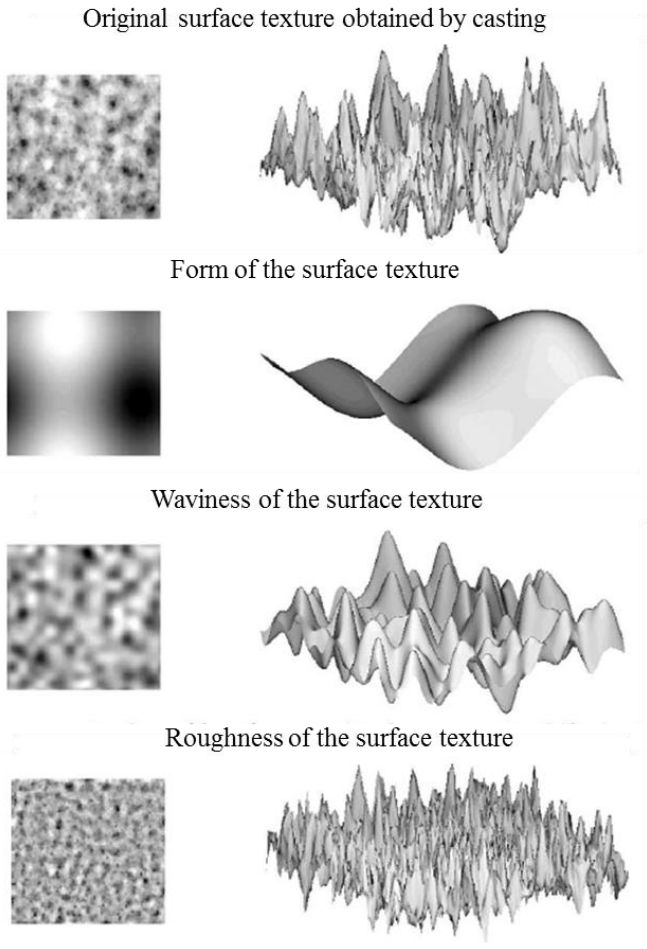


Figure 1.4 Form, waviness and roughness decomposition of a surface texture using frequency normalized wavelet transform (FNWT) and applying Daubechies wavelet of order 20 [65].

1.3.7 Slope analysis techniques

Mid-spatial frequencies are sometimes called ripple, waviness or slope errors. ISO 10110-8, the International Organization for Standardization of optics and photonics drawing standards, has several notations for specifying MSF [74, 75]. Among those, the simplest one is RMS waviness, or W_q. The RMS waviness though is not a suitable metric for specifying MSF. A surface produced by small tool computer controlled polishing with smooth ripples, a cusp-shaped surface produced by diamond turning and a random isotropic surface might have the same RMS waviness but have different effects on performance. The simulated effect of cusp-shaped waves *vs*. sinusoidal waves on optical performance is shown in chapter four. The maximum slope error $\delta_{\max}$, the maximum deviation amplitude $A_{\max}$ and the maximum sampling length $L_{\max}$ are other parameters defined in ISO 10110 to calculate the slope error and are illustrated in Figure 1.5 [74, 75]. These values are chosen relative to the surface that best fits to the nominal one. The maximum slope error $\delta_{\max}$ should be measured at maximum sampling length $L_{\max}$. Surface slope is the change in OPD *vs.* small transverse shift and is highly relevant to MTF. It also correlates well with the optical performance at low spatial frequencies [76].

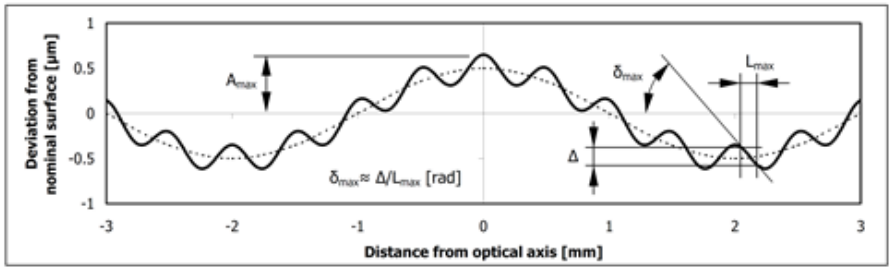


Figure 1.5 The maximum slope error $\delta_{\max}$, maximum deviation error $A_{\max}$ and maximum sampling length $L_{\max}$ are ISO 10110 requirements for mid-spatial frequency (waviness) content on a surface [75].

1.4 Navigating this dissertation

So far, we have described mid-spatial frequencies, and why it is important to specify and characterize them. In section 1.3, we have listed different mathematical methods for characterizing MSF such as PSD, structure function, Zernike and Forbes polynomials. In chapter two, using published discussions from the literature along with some data analysis on measured surfaces, we compare pros and cons of each analysis for MSF texture based on chosen evaluation criteria. Our evaluation criteria are the amount of pre-processing, the uniqueness of the results (related to orthogonality), connection to fabrication choices, dependence on aperture geometry, generalizable to full frequency range and connection to optical performance. The focus of this dissertation is on investigating a Zernike polynomial representation for quantifying the mid-spatial frequency content in the height maps. In chapter three, we will illustrate how a Zernike polynomial representation captures certain characteristics of the mid-spatial frequency texture such as spatial length scale, the magnitude of length scale, and type of MSF texture. In section 3.2, we show how recursive formulas are essential in the generation of extreme orders of Zernike polynomials. After enabling the generation of these orders, in section 3.3, we build up an intuition for interpretation of fit coefficient maps by analysis of synthetic data and extend our analysis to real measured surfaces with MSF content. Finally, we discuss limits of Zernike analysis for finite datasets and convergence of RMS fit residuals. In chapter four, we introduce optical performance metrics for MSF such as Strehl ratio and modulation transfer function. We will show examples of optical performance degradation in presence of synthetic and real measured MSF textures. The effect of PV, incident wavelength, refractive index and type of MSF texture on optical performance will be discussed. An ideal analysis method is

one that enables data reduction to metrics that have a clear connection to optical performance so that designers can incorporate this metric in their tolerancing. In chapter five, we extend Zernike polynomial analysis to the prediction of optical performance based on the decomposition of the surface into bands of Zernike polynomials. We will show the relationship between the Strehl ratio and MTF of the whole surface to the Strehl ratio and MTF of individual bands. We will discuss limits of validity of our approximations. In chapter six, after summarizing conclusions of this work, we will list assumptions and range of validity of our optical performance analysis in the Rayleigh-Sommerfeld regime. Finally, in section 6.3, we have included suggestions for future research in extending Zernike analysis to combine data captured from multiple instruments.

CHAPTER 2 : CRITERIA TO BE CONSIDERED FOR SPECIFICATION AND CHARACTERIZATION METHODOLOGY

In this chapter, the evaluation criteria for characterization of mid-spatial frequency will be described and the mathematical methods will be evaluated based on each criterion. Our evaluation criteria are the amount of pre-processing, the uniqueness of the result (related to orthogonality), connection to fabrication choices, dependence on aperture geometry, generalization ability to full frequency range, connection to optical performance and intuitiveness. Each of these criteria will be described and the mathematical methods will be evaluated based on those criteria.

2.1 Amount of pre-processing

Preprocessing is processing of the raw data and preparing it for performing the actual characterization. We will have the most robust and unambiguous analysis output if this step is standardized before the actual process takes place. The results of the analysis will be more repeatable and the same for different metrologists. The preprocessing choices mainly include filtering, cropping, windowing and padding the data.

2.1.1 Pre-processing for PSD analysis

To prepare the data for PSD analysis the first thing we need to do is to remove the low order surface figure and rigid body alignment terms. Removal of these surface errors is referred to as de-trending [42]. Then, since PSD is a Fourier-based analysis, we need to set edges of the raw data to zero by applying a 2D window function. If we do not apply the window function, the PSD will have artifacts induced by the edge discontinuities aligned

along the x and y directions. By applying a window, PSD will provide a much better estimate of the surface roughness. Windowing or apodization is a well-known technique in signal processing [77, 78]. Window functions are usually smooth and positive curves that have zero values outside a chosen interval. Uniform window means no apodization is applied. Han, Hamming and Blackman windows are commonly used in the calculation of power spectral density and in signal processing [77-79]. These windows are within $\cos^{\alpha} x$ family that are described by Eqn. 2.1. In this family of windows, changing the parameter α generates different windows [77].

$$f[n] = \cos^{\alpha}\left(\frac{\pi n}{N}\right), \quad 0 \le |n| \le \frac{N}{2} \tag{2.1}$$

In Figure 2.1, P. Z. Takacs shows the un-windowed and windowed area PSDs of the surface data. As can be seen in Figure 2.1 (b) the edge discontinuity in un-windowed data introduces spurious power along the x and y frequency axes [42]. This artificial oscillation is due to Gibbs phenomenon and the inherent difficulty of approximating a discontinuous function by a finite series of harmonic functions.

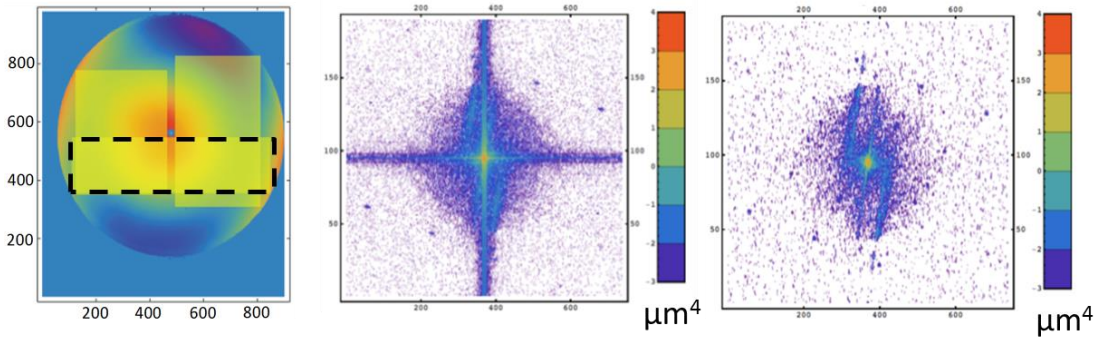


Figure 2.1 (a) Wavefront map of a lens from Fizeau interferometer. Area PSD of the dashed selection in the map after de-trending 4th order polynomial, (b) when there is no window and (c) with an application of a 2D Blackman window [42].

There are different window functions available and depending on the choice of the window the statistics derived from the PSD vary [42]. Figures 2.2 and 2.3 show that the type of window can affect the results by an order of magnitude. Figure 2.2 (a) is the residual surface of SiC sphere in Figure 1.1 (c) after removing the first 36 terms of Zernike polynomials in the standard set (shown in Table 3.1). We have de-trended our data using Zernike polynomials due to its circular aperture. P. Z. Takacs has removed least square fits of 4^{th} order polynomial in x and y to de-trend the rectangular data [42]. Please refer to section 3.1 for a detailed description of Zernike polynomials and Table 3.1 for mathematical representation of the first 36 terms in the standard set. In Figure 2.2, a linear PSD calculation is carried out to a linear profile perpendicular to the direction of the raster patterns on the surface. Figure 2.3 is the non-directional PSD of the same data with a choice of different windows. Non-directional PSD is calculated by summing the 2D PSD for every frequency in all directions. This is performed by adding all the PSD values on a circle that have the same frequency (Eqn. 2.2) [80].

$$NPSD(\left|\vec{f}\right|) = \int_0^{2\pi} PSD(\vec{f}\,')\left|\vec{f}\,'\right| d\theta_{f'}, \tag{2.2}$$

where f is the frequency and $\theta_{f'}$ is the azimuthal angle at each frequency. Figures 2.2 and 2.3 show that depending on the type of window, the PSD results can change by an order of magnitude for both directional and non-directional PSDs. In directional PSD in Figure 2.2 (c), the PSD value for the spatial period of 0.7 mm ($1.4\ 1/mm$) is 0.01 for a case of the uniform window. The PSD value at the same spatial period is 0.1 when we apply a Hanning window. In non-directional PSD, in addition to PSD peaks with different values at 0.7 mm ($1.4\ 1/mm$) and 1.6 mm ($0.6\ 1/mm$), the overall PSD with the uniform window is smoother

and higher at high spatial frequencies compared to PSD when Blackman and Hanning windows are applied.

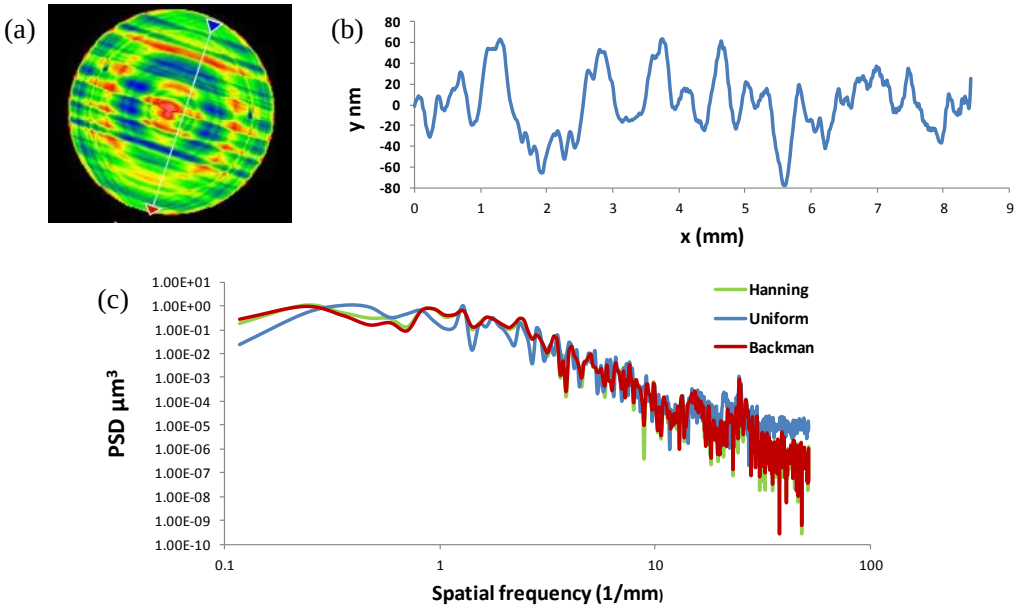


Figure 2.2 (a) SiC sphere residual map after de-trending the first 36 terms Zernike polynomials in the standard set. (b) Selected surface profile perpendicular to raster patterns. (c) PSD of the linear profile with the application of different windows.

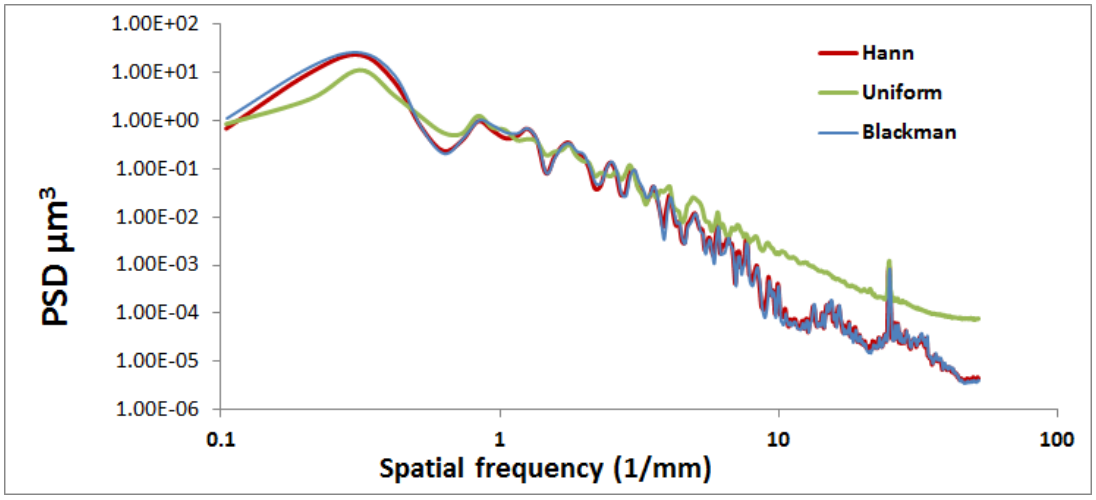
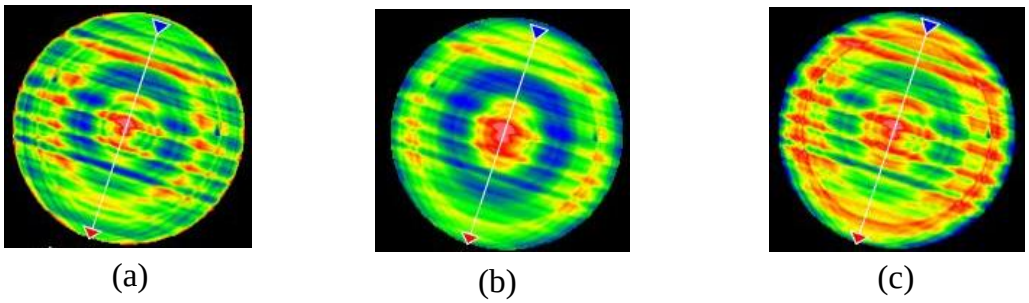


Figure 2.3 Non-directional PSD of the de-trended SiC sphere (map on Figure 2.2 (a)) with the application of different windows.

Figure 2.4 shows how different de-trending choices affect results of the PSD. We have obtained linear PSD of the profile shown in Figure 2.2 on surface data of SiC sphere using the MetroProx software. At the spatial period of 2.1 mm ($0.47\ 1/mm$), the PSD value for different de-trending choices is different by orders of magnitude. For example, for removal of seven orders of Zernike polynomials in the standard set, (the first 36 terms in Table 3.1, i.e. terms up to $\rho^7 \sin(7\theta)$), the PSD value is 0.22 μm^3 and for removal of four orders of the Fringe set, the PSD magnitude is 0.077 μm^3. The Fringe set has a different ordering than the standard set. Table 2.1 shows the first four orders of Zernike polynomials in the Fringe set. For spatial periods larger than 2.85 mm (frequencies smaller than $0.35\ 1/mm$), the PSD of Fourier filtered data is the same as PSD of data after removal of 7^{th} orders of the standard set. At the spatial period of 2.1 mm ($0.47\ 1/mm$) PSD of Fourier filtered data is closer to the PSD of the data where we have removed 4^{th} orders of the Fringe set. The PSD of all the data for all three de-trending choices are the same for spatial periods smaller than 1.2 mm (frequencies larger than $0.83\ 1/mm$) which is not surprising since de-trending affects low spatial frequencies and keeps the high-frequency content unchanged. The analysis for Figure 2.2 to 2.4 was performed using MetroProX (Mx) software (v6.3.0.4). So far, we observed examples of how choices of windowing and de-trending can change the result of PSD calculation which can cause an economic loss if a good part does not meet the specification defined by the customer or designer or a bad part is shipped as good.



(a) (b) (c)

(d)

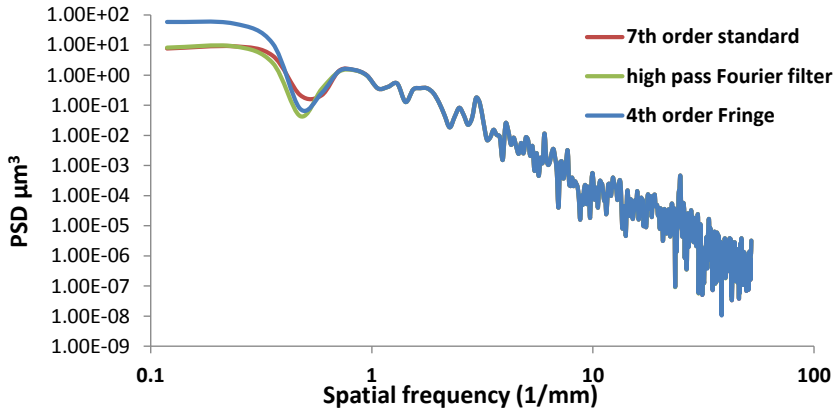


Figure 2.4 SiC sphere residual map after removing (a) Seven orders of Zernike polynomials in standard set, (b) Four orders of Zernike polynomials in Fringe set, and (c) spatial periods greater than 3 mm using Fourier filter (d) PSD of the linear profile with the application of different de-trending choices in part a-c.

Table 2.1 Zernike polynomials in polar coordinates (the Fringe set) [29, 47].

ZFR	n_z	m	Sets	RMS	Orthogonal polynomials	Aberration names
0	0	0	Z_0^0	$\sqrt{2}$	1	piston
1	1	1	Z_1^1	2	$\rho\cos\theta$	x-tilt
2	1	-1	Z_1^{-1}	2	$\rho\sin\theta$	y-tilt
3	2	0	Z_2^0	$\sqrt{3}$	$-1+2\rho^2$	focus
4	2	2	Z_2^2	$\sqrt{6}$	$\rho^2\cos(2\theta)$	Astigmatism 0 &
5	2	-2	Z_2^{-2}	$\sqrt{6}$	$\rho^2\sin(2\theta)$	Astigmatism 45 &
6	3	1	Z_3^1	$2\sqrt{2}$	$(-2\rho+3\rho^3)\cos(\theta)$	Coma & x-tilt
7	3	-1	Z_3^{-1}	$2\sqrt{2}$	$(-2\rho+3\rho^3)\sin(\theta)$	Coma & y-tilt
8	4	0	Z_4^0	$\sqrt{5}$	$1-6\rho^2+6\rho^4$	Spherical & focus

2.1.2 Pre-processing for Zernike polynomial analysis

Zernike polynomials are orthogonal over a circular aperture and their fit coefficients show the amount of each aberration term present on the surface. To analyze the circular aperture data with Zernike polynomials, there is usually no need to pre-process the data. Gram-Schmidt methods can be applied to make Zernike polynomials orthogonal over non-circular apertures and for discrete data sets [81]. For rectangular or square apertures, using Chebyshev and Legendre polynomials is also recommended [82, 83]. More details on Zernike analysis for different aperture geometries are discussed in section 2.4.

2.1.3 Pre-processing for Forbes polynomial analysis

In order to fit data to Forbes Q-polynomials, removing the high-frequency content of surface data is required since the number of selected data points from the surface should match the number of fitted coefficients to avoid aliasing. Additionally, using a non-uniform grid will minimize the effects of aliasing. Figure 2.5 shows the sample locations for describing an example freeform surface from the patent literature (discussed in Section 2.2 of reference [51]) using $Q_{n_f}^m$ polynomials [49]. Note that so far, Forbes analysis has been shown for circular apertures. More general and less efficient versions of Forbes polynomials exist that can fit to non-circular aperture shapes [84].

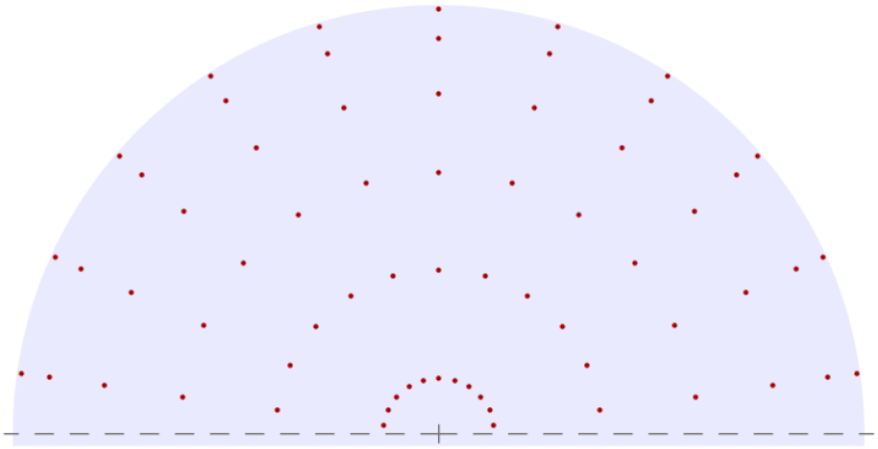


Figure 2.5 Selected data points on six rings and eleven spokes for describing an example freeform surface (discussed in Section 2.2 of reference [51]) [49].

2.1.4 Pre-processing for structure function analysis

Structure function is a direct calculation that does not involve additional processing such as windowing, but it is sensitive to background removal (excluding piston) because it is the expectation of the squared height difference as a function of separation. As a result, for surfaces containing low order form errors, we will have a non-zero SF for small separations. Figure 2.10 in reference [64] shows the SF result for a sinusoidal profile with both low and high-frequency content. Extracting useful information from the SF in presence of low order errors is difficult [64]. Therefore, L. He recommends removal of low-order polynomials (or low-frequency errors) before the structure function calculation for characterizing mid-spatial frequency.

2.1.5 Pre-processing for ACF analysis

This calculation does not involve windowing, but ACF is sensitive to background removal choices including piston and tilt due to its definition. Again, removal of piston and tilt for flat surfaces and piston, tilt, and power for spherical surfaces should be considered

before the ACF calculation to avoid this sensitivity. Additionally, to calculate ACF, profiles should be measured from the same mean plane to avoid singularities [57].

2.2 Uniqueness of the result (related to orthogonality)

If a mathematical representation is orthogonal, the results are unique regardless of whether or not the whole surface is processed or an arbitrary summation of bands within the surface is processed. In orthogonal basis sets, the fit coefficients are unique and impendent of the particular subset of the basis elements used for the fit. The benefits of having an orthogonal representation are that the analysis results are not impacted by user-defined choices. For example, the results of an orthogonal representation will not be impacted by choices of the different spatial filter in the pre-processing step.

2.2.1 Orthogonality of PSD analysis

Harmonic functions are the basis set for PSD analysis that are orthogonal over an infinite domain. Real datasets though are usually discrete and truncated. From Plancherel's theorem, we can calculate the energy distribution over a frequency range for a filtered signal of interest. Plancherel's theorem states that the integral of a function's squared modulus is equal to the integral of the squared modulus of its frequency spectrum [85]. Since these harmonic functions are orthogonal over an infinite domain, the PSD calculation of a finite dataset is not an orthogonal representation, which is why the results are not unique – they depend on the use of a window, the type of window and de-trending choices. The results can vary by an order of magnitude and this will impact the RMS extracted from the PSD (Eqn. 1.4). Examples of pre-processing choices and their effect on the PSD results are shown in the previous section (section 2.1).

2.2.2 Orthogonality of Zernike polynomials

Zernike polynomials are an orthogonal basis set over a continuous unit circle. These polynomials are not orthogonal over discrete datasets. The problem of orthogonality associated with the discrete dataset is more obvious when fitting to large orders of these polynomials. If we use the orthonormal Zernike functions to represent the surface error, the coefficient of each mode (polynomial of order *n* and *m*) is the contribution of that mode to the RMS wavefront of the dataset and the fitted coefficients are unique. Moreover, the RMS of a surface constructed from a sum of orthonormal Zernike polynomials can be obtained when the coefficients are added in quadrature (root sum square of the coefficients). When fitting to large orders of Zernike polynomials using discrete datasets and conventional fitting methods, the approximation to orthogonality degrades. Kintner has shown recurrence relations for Zernike polynomials through their relation to Jacobi polynomials [86]. G. Forbes has used this recurrence relation to numerically generate large orders of Zernike polynomials [87]. Limits of Zernike polynomials for a description of mid-spatial frequency and concerns about orthogonality of these polynomials at very large orders will be described in chapter three.

2.2.3 Orthogonality of Forbes polynomials

Forbes polynomials are orthogonal in slope. Fitting or projection onto $Q_{n_f}^m$ polynomials gives a direct representation of the RMS gradient. The RMS gradient is connected to the ray spot diagram [49]. The RMS slope is related to transverse ray error described in section 2.6. Spot diagram shows the location of transverse ray errors. The RMS spot size is a common image quality metric and is the center of gravity of transverse ray errors [14, 28]. The approximation to orthogonality degrades when fitting discrete datasets with uniform

pixel spacing to extreme orders of these polynomials.

2.2.4 Orthogonality of structure function

The discussion of orthogonality for structure function is irrelevant since orthogonality is a characteristic of basis set and structure function is a calculation. A basis set is orthogonal when the inner product of two basis elements is defined by the delta function. L. He has compared the SF of Zernike polynomials individually with the SF of a surface described by a combination of multiple Zernike polynomials. The total linear SF of the sum of Zernike polynomials is the sum of the individual linear SFs of each of the Zernike polynomials with different azimuthal frequencies [64]. Therefore, the linear SF is a linear operator over Zernike polynomials. This property is not valid for area SF. Since area SF is not a linear operator, combining and representing data from multiple instruments requires additional steps besides plotting the data on the same scale [59].

2.2.5 Orthogonality of ACF

Like the structure function, the discussion of orthogonality for autocorrelation is irrelevant. Furthermore, the total ACF of the sum of Zernike polynomials is not the sum of the individual ACF of each of the Zernike polynomials. The ACF also depends on the mean plane.

2.3 Connection to fabrication choices

This feature of the analysis is about connecting the result of the analysis to a fabrication detail and how the analysis can be used for optical specification. This connection could also be made between the analyses of data within a specific band. An ideal representation will allow the fabricator to determine the length scale, the orientation and type of the texture (e.g. raster, spokes or ring-like texture), as well as the magnitude of each length scale.

2.3.1 PSD and connection to fabrication

It is possible to set a maximum allowable PSD for a specific frequency range or limited spatial bandwidth. The PSD of the fabricated element should not exceed that line. The area under the linear PSD curve within a certain frequency range is the RMS of that frequency range which is a useful parameter that can be considered in the specification. For example, in reference [39] one can see a list of PSD related parameters that the national ignition facility (NIF) used for glass optics. These parameters include PV, RMS gradient, r_q (RMS height) and the amplitude parameter of the PSD not-to-exceed line. In Figure 2.6, the mid-spatial frequency region is categorized in PSD-2 (spatial length scales between 2.5 mm to 0.12mm). For this spatial bandwidth, the RMS should not exceed 1.6 nm and the amplitude A should not be above $1 nm^2 mm$. A peak in the PSD that is above the specification line can be converted into length scale and the fabricator knows what length scale on the surface needs to be removed. PSD is effective at highlighting specific spatial frequencies that result from the machining process. This can be useful in providing feedback to improve fabrication and change fabrication parameters such that the part will meet the specification. An example of this is changing the programmed tool path in diamond turning machine or choice of the tool footprint in the magnetorheological finishing (MRF) process. The PSD analysis can quantify the surface so that periodic structure in the part can be minimized during manufacture. To speculate on the effectiveness of PSD in retrieving surface finish information (length scale, type, magnitude) we have calculated the PSD of a linear trace from the surfaces shown in Figures 2.7 (a) and 2.7 (b). Figure 2.7 (a) is a linear sinusoidal pattern with 25 cycles across aperture and PV of 2 μm and Figure 2.7 (b) consists of 25 sinusoidal concentric rings with PV of 2 μm. The linear PSDs from both surfaces are

identical and is shown in Figure 2.7 (c). We see there is a peak at $f_x = 12.5\frac{1}{mm}$ that is equal to the spatial period of $\Lambda = 0.08mm$. PSD provides information on the magnitude of length scales through the calculation of RMS surface height and integral under the PSD curve. The RMS of the surface for surfaces in Figures 2.7 (a) and (b) are 0.71 µm that matches with the theoretical RMS of a sinusoidal surface with a PV of 2 µm. So far, we observed that PSD provides information on the length scale and magnitude of texture.

Extracting information about length scale, magnitude, and type of mid-spatial frequencies involves slicing through the two-dimensional PSD or integrating the two-dimensional PSD into a one-dimensional PSD. In order to do so, prior insight into the direction of MSF texture is required. For example, selecting a horizontal slice of two-dimensional PSD is suitable for linear MSF texture but will not provide information on the spokes patterns present in the surface. The fact that PSD of both surfaces in Figures 2.7 (a) and (b) are the same confirms that PSD is not an ideal representation for identifying the type of the texture, especially for anisotropic texture.

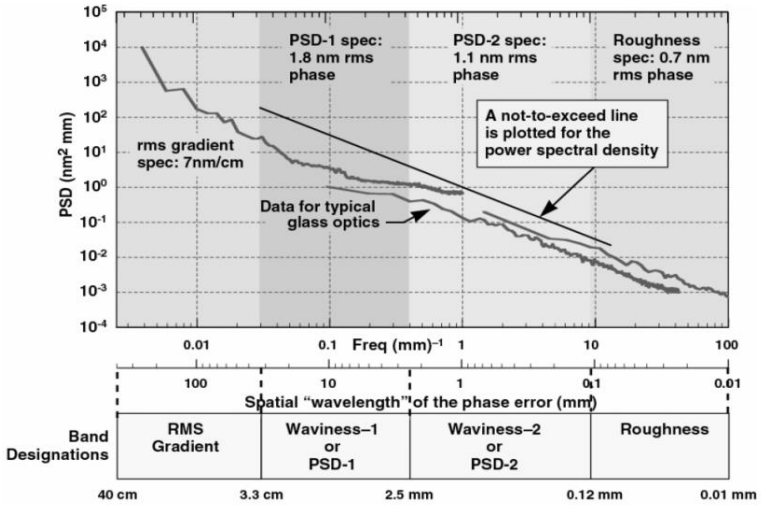


Figure 2.6 PSD specification for an NIF optic [39].

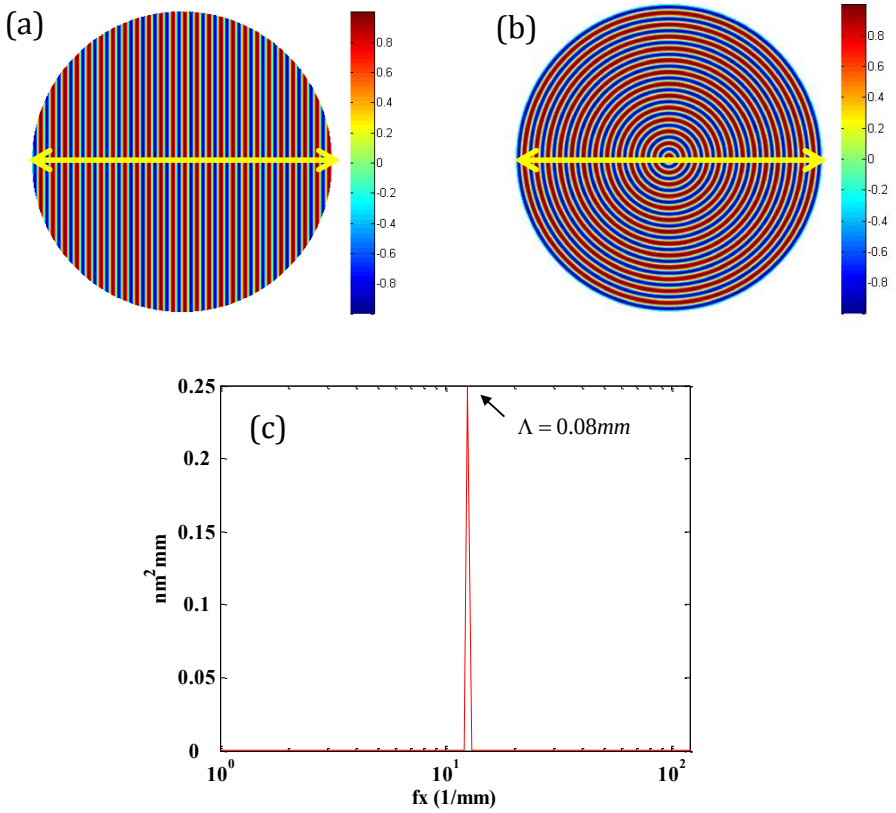


Figure 2.7 (a) Height map of sinusoidal raster pattern with 25 cycles within aperture and PV of 2 µm. (b) Radial sinusoidal pattern with 25 cycles across aperture and PV of 2 µm. (c) PSD of the linear trace from data in (a) and (b).

2.3.2 Zernike analysis and connection to fabrication

Optical surface or wavefront specifications can include descriptions in terms of Zernike polynomials. Commercial software in modern interferometers frequently provide an interface to fit the wavefront or surface figure error to a set of Zernike polynomials. For example, the form application in MetroProX 6.3.0.4 software provides options to fit and remove Zernike polynomials. The software provides 37 terms of Zernike Fringe set and 91 terms of Zernike polynomials in the standard set. Opticians frequently use the Zernike polynomials both in testing, analysis, and diagnosis of optical components. C. J. Evans *et al.* have shown visual representations of Zernike coefficients in a form that helps optical

fabricators decide how to improve their process [88]. For example, the coefficients a_n^m and a_n^{-m} for $m \neq 0$ are added in quadrature to provide the aberrations of those orders independent of the angular orientation [88]. In the next chapter, we will show how large orders of Zernike polynomials can be used to capture and quantify the MSF texture. In section 3.3.1, we will show that location and orientation of the peak of the fit coefficients and RSS (Root Sum Square) of the Zernike fit coefficients can be used to retrieve information about length scale, orientation (type) and magnitude of the MSF texture. We will also showcase filtering of height maps based on symmetry property of Zernike polynomials. This unique filtering capability has potential in identifying and separating different sources of error during the fabrication process. For example, the tool tip could leave raster-like texture on the part and errors in the motor could lead to spokes fingerprint.

2.3.3 Forbes analysis and connection to fabrication

Forbes polynomials can be used to specify nominal shape as well as mid-spatial frequency components in freeform surfaces [49]. They can also show signatures of different tooling marks on the surface in MSF region. MSF patterns can be raster-like or rotationally invariant texture. Forbes analysis provides RMS gradient of the surface that may be used for specification. Q-polynomial representation and Zernike polynomials behave the same in retrieving information about the length scale and type of the MSF texture. The location of the peak of coefficients happens at specific Cartesian orders t that is related to the number of cycles across aperture (length scale) through $t \approx \pi C$ where C is the number of cycles across aperture [49]. The Cartesian order t is indicative of degrees of freedom when describing a surface using Zernike or Forbes polynomials, *i.e.* it is the maximum power of u in Eqn. 1.7 [49]. Orientation and location of the peak of Q-

spectra provide information on the type of MSF texture. E.g. the raster MSF texture will show a diagonal peak in the Q-spectra. The sum of the squares of the coefficients along t is the weighted mean square gradient of the fitted Q-polynomials. *i.e.* the RSS of the Q-spectra along the Cartesian order t provides the RMS gradient [49].

2.3.4 Structure function and connection to fabrication

Structure function can have potential application in the specification of general optical surfaces. A connection has been made between SF and RMS gradient of the surface in [64].

$$SF(r \to 0) = \frac{r^2}{2}(\nabla z)^2_{RMS}. \tag{2.3}$$

where $r = |P_1 - P_2|$ and p_1 and p_2 are two different points on the surface. For fractal surfaces (self-affine surfaces), the SF is a function of fractal parameters, the fractal dimension *D* and the topothesy Λ.

$$SF(\tau) = \Lambda^{2D-2}\tau^{2(2-D)}. \tag{2.4}$$

Topothesy is indicative of the magnitude of the fractal dimension. The fractal dimension and topothesy are determined by measuring the slope and intercept of a logarithmic plot of SF [56]. Machined surfaces with multiple length scales, produced by more than one finishing process, are called multifractal and their SF has two or more straight lines of different slopes meeting at a more or less sharp discontinuity [56]. For a stationary surface, the SF does not have advantages over ACF. To speculate the effectiveness of structure function in providing information on length scale, magnitude, and direction of MSF texture, it is helpful to look at the SF of a sinusoidal pattern. L. He has simulated the linear SF of a sinusoidal profile and the SF is also a periodic function. The periodic minima of the SF are separated by the spatial period of the height profile since the SF is zero when the separation

values are exact multiples of the spatial wavelength. However, the distance between the consecutive maxima is not exactly equal to the spatial wavelength of the profile since SF is the average of the squared height differences. There is also a relationship between the PSD and linear SF for stationary surfaces:

$$PSD(f) = \Delta\tau.\mathcal{F}[ACV(\tau)] = \Delta\tau.\mathcal{F}[\sigma^2 - \frac{1}{2}SF(\tau)]. \quad (2.5)$$

This relationship can be useful in retrieving the amplitude of a specific length scale. The area SF has information on the anisotropy of the surface and the direction of MSF (sometimes referred to as ripples). The direction of MSF texture in the surface data is the same as the direction of oscillations in the area ASF as shown in Figures 10 and 11 of reference [58].

2.3.5 ACF and connection to fabrication

Autocorrelation length is the distance at which the ACF drops to $1/e$ of its initial value. The autocorrelation length is a useful metric in capturing hidden periodicity in random surfaces. For sinusoidal structures, the ACF has the same spatial period as the surface profile. The value of the ACV at the origin is the variance of the surface (σ^2) or R_q^2. Linear ACF can be used to analyze mechanically ground surfaces and fabrication processes. Whitehouse indicates that information like the randomness of the process, the efficiency of the cutting, the hardness of the workpiece material and subsurface damage can be revealed from the ACF [61]. Based on the Wiener–Khinchin theorem for a well behaved random stationary process, there is a relationship between the PSD and ACV [89]. The PSD is the product of $\Delta\tau$ and Fourier transform of the autocovariance function.

$$PSD(f) = \Delta\tau.\mathcal{F}[ACV(\tau)]. \quad (2.6)$$

Therefore, we can also indirectly calculate the length scale and magnitude of the length scale from the PSD obtained from ACV for a random and stationary process.

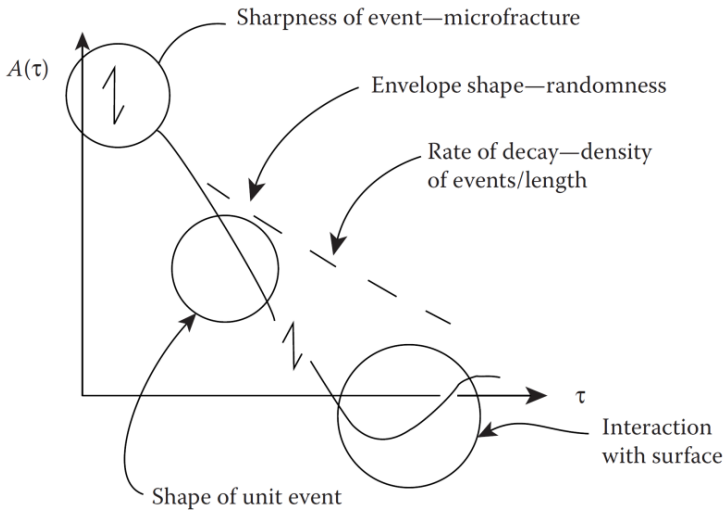


Figure 2.8 Relationship between ACF shape and grinding [61].

2.4 Dependence on aperture geometry

An ideal analysis method for freeform optics and MSF would be a method that could analyze any arbitrary aperture geometry efficiently and effectively. It is a matter of discussion whether or not the analysis will be correct and effective if applied to an arbitrary aperture shape. In this section, we will discuss the consequences of applying the analysis to a dataset within an aperture for which the calculation is not well suited.

2.4.1 Arbitrary aperture for PSD analysis

Since PSD is a Fourier-based analysis and makes use of DFT, calculation of area PSD only applies to rectangular datasets. For circular data sets, one has to set the corners to zero or use inscribed squares of rectangles [42].Setting corners to zero will have an effect on the overall RMS values and will reduce the magnitude for higher frequencies. By using

inscribed squares or rectangles, we will lose some of the real data points on the surface. An alternate approach to avoid this problem is to calculate the PSD based on radial slices through the circular dataset. P. Takacs has integrated over all azimuthal angles in 2D PSD data to generate an equivalent radial PSD [42].

2.4.2 Arbitrary aperture for Zernike analysis

Zernike polynomials can be defined for circular, annular, and elliptical apertures [81, 90]. W. Swantner and W. Chow have shown that Gram-Schmidt method can be used not only on circular or annular apertures but also on any aperture shape [81]- see also ISO 14999-2. For a basis set of arbitrary shape defined by $\{V_1.....V_{n-1}\}$, the Gram-Schmidt method is represented by

$$V_n' = Z_n + \sum_{m=1}^{n-1} D_{nm}' V_m, \tag{2.7}$$

where $\{V_1'.....V_{n-1}'\}$ are the new set of orthogonal but not orthonormal basis of arbitrary shape, $\{Z_1.....Z_{n-1}\}$ are Zernike circular polynomials, and $V_n = V_n'/\|V'\|$ and D_{nm}' can be calculated by $D_{nm}' = -\langle Z_n, V_m \rangle$ [91].

If Zernike circle polynomials are used for non-circular apertures, the orthogonality condition is violated and the coefficients can no longer be added in quadrature to provide the RMS wavefront. In order to make them orthogonal again, the Gram-Schmidt orthogonalization described above should be applied.

2.4.3 Arbitrary aperture for Forbes analysis

Forbes Q-polynomials can be applied to circular aperture shapes very efficiently where they provide almost one fitted coefficient for every sample point evaluated within the

circular aperture. [49]. For non-circular apertures, there are modifications based on least-squares methods that can be applied to Forbes polynomials for non-circular apertures which make the computation less efficient [49].

2.4.4 Arbitrary aperture for structure function

Structure function is a direct calculation without a basis set and is computationally correct for any arbitrary aperture shape without extra processing and this is one of the main advantages of this analysis [64].

2.4.5 Arbitrary aperture for ACF analysis:

ACF is a direct calculation and can be applied to any type of aperture geometry [62].

2.5 Generalizable to full frequency range

This characteristic of the characterization method is whether or not the analysis can be applied to multiple length scales of the surface in the regions of the form (figure), mid-spatial frequency and finish (roughness). Not all analysis methods are capable of characterizing all length scales of the data. For example, PSD is not capable of characterizing low order form errors. The presence of low order errors on the profile will impact the analysis in other frequency regions by introducing frequency leakage. Another aspect of this feature is representing combined data from multiple instruments with different length scales. Usually, one instrument is not capable of capturing all the surface characteristics all the way from form to waviness and finish. Properly combining the data in an analysis of different mathematical representations has unique issues for each mathematical method that we will discuss in this section.

2.5.1 PSD analysis for full frequency range

Different linear PSDs from measurements of multiple instruments can be combined

and plotted on the same representation. However, since PSD uses fast Fourier transform (FFT) method, for the analysis to be correct, low order terms should be removed. Therefore, PSD analysis is not applicable for representation of form errors. E. Sidick has shown the effect of low-order figure error on PSD results. He has shown that it is necessary to remove all of the low-order figure errors from the surface height map before carrying out the PSD calculation since they can leak into all frequency components in PSD domain and make the results inaccurate [92].

2.5.2 Zernike analysis for full frequency range

Theoretically, any circular surface can be expanded based on an infinite number of Zernike polynomials which can cover all spatial frequencies on a surface. Large orders of Zernike polynomials can be generated using recurrence relation to avoid numerical round-off errors associated with explicit equations. As the order of Zernike polynomials increases, denser sampling is needed to capture all the oscillations within the polynomial. Therefore, the surface data has to have a minimum number of pixels to be able to correctly fit to large orders of Zernike polynomials. I. Kaya *et al.* have shown that in order to fit to a Zernike polynomial set $9k^2$ number of samples is needed where k is the largest order of those polynomials [93]. In interferometers, the instrument transfer function (ITF) shows a decreasing response as surface frequencies approach Nyquist. Contact profilers, microscopes, CCDs and almost all other surface measuring instruments provide pixel limited datasets that have a cut-off frequency associated with them. In order to cover a larger range of frequencies from figure to finish, measurements from different instruments should be combined. Thunen *et al.* [94] and Chow and Lawrence [95] have developed analytical methods based on Zernike polynomials for reconstructing full aperture

wavefront from sub-aperture measurements.

2.5.3 Forbes analysis for full frequency range

Forbes Q-polynomials can be applied to all length scales available in the data. Fitting can be carried out to extreme orders as long as the number of fit coefficients and number of the sampling points match [84]. This capability makes Q-polynomials suitable for MSF characterization on freeform surfaces [49]. Forbes Q-polynomials also have the same problem as Zernike polynomials when fitting to surface data of limited size.

2.5.4 Structure function analysis for full frequency range

Structure function is not an orthogonal representation and combining data from multiple instruments is challenging especially for non-stationary surfaces such as ones with MSF. This is due to the fact that a surface with form error will have a non-zero SF for small separations. Usually, the tilt is the main form error that is unknown in sub-aperture measurements. Therefore, tilt in the raw sub-aperture data must be replaced by the correct tilt in the full aperture data [59]. L. He *et al.* have gathered data from multiple instruments and presented SF interpretation of the combined data [64]. L. He recommends removing low-order form errors to extract useful information from the structure function [64].

2.5.5 ACF analysis for full frequency range

Since ACF is not an orthogonal representation, we expect that combining data from different instruments in one representation will be non- intuitive. Like SF, representing data from multiple instruments requires additional steps such as correcting the tilt in the sub-aperture data.

2.6 Connection to optical performance

This feature of the analysis model enables us to draw some connection between features

of the surface and the optical performance. The performance metrics are application specific and include performance in illumination or imaging systems. Common image quality metrics are RMS spot size, encircled energy and Strehl ratio. MTF can also be used in quantifying image performance. These metrics and connection of MSF characteristics to optical performance will be described in detail in chapters four and five.

2.6.1 PSD and connection to optical performance

J. E. Harvey *et al*. have shown how different regions of the PSD curve have different scattering characteristics [96]. Low spatial figure errors scatter the energy from the image core to the first few diffraction orders reducing the Strehl ratio without significantly broadening the central image peak. Surface finish or high spatial frequency errors scatter the light from the central peak and result in reducing the peak but without significantly broadening the image core and changing the resolution. It is the mid-spatial frequency region which broadens the image core and significantly reduces the resolution. Therefore, he argues the surface PSD is particularly a relevant statistical quantity relating scattering behavior to surface topographic features [96]. The RMS of the whole surface (that includes all length scales) is not an appropriate metric for specification since different length scales have a different impact on the point spread function and MTF. For optical performance prediction, form and mid-spatial frequency regions should be separated since the MTF behavior for low order terms and MSF region with the same RMS is different [24]. During the de-trending pre-processing step for PSD calculation, most of the low order form errors are removed in order for the PSD results to be accurate for higher frequency regions. Therefore, PSD analysis cannot be connected to image quality and optical performance for low order form errors and aberration terms.

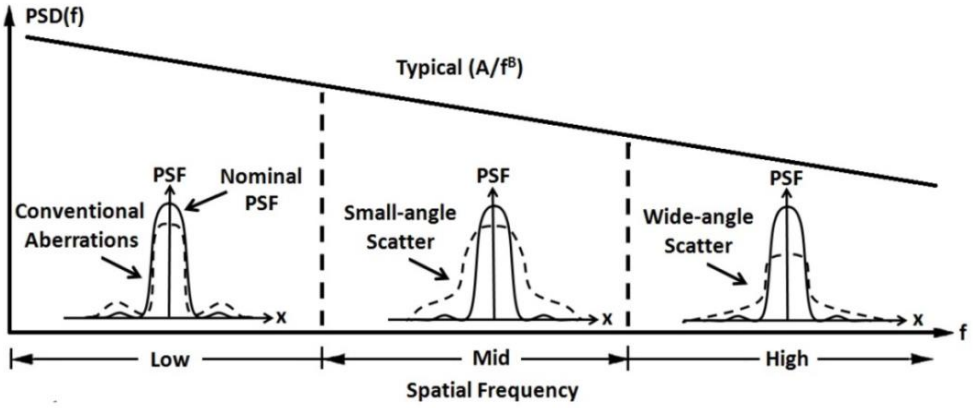


Figure 2.9 Effect of different spatial frequencies on the point spread function (PSF) [96].

2.6.2 Zernike polynomials and connection to optical performance

The metrics of image quality such as RMS, Strehl ratio and the area under the MTF curve can be extracted from the fit coefficients of Zernike polynomials [14, 29]. This connection has been made between low order Zernike polynomials (form errors). The low order Zernike polynomials are directly related to first order and third order aberration terms (important in many imaging systems) and can be calculated for any image point (Eqn. 1.2). For small aberrations, the Strehl ratio is equal to $S = \exp(-k^2\sigma^2)$ where k is the wavenumber and σ is the RSS of the Zernike fit coefficients for an orthonormal representation [97].We will discuss the connection between optical performance and large order Zernike polynomials that describe MSF texture in chapters four and five.

2.6.3 Forbes polynomials and connection to optical performance

The RSS of the fit coefficients of Q-polynomials along Cartesian order t is equal to the wavefront's RMS slope. The RMS slope is related to image quality metrics such as RMS spot size [28]. The slope of the wavefront is related to transverse ray error through Eqn. 2.8.

$$\begin{cases} \varepsilon_Y(x_p, y_p) = -\dfrac{R}{r_p}\dfrac{\partial W(x_p, y_p)}{\partial y_p} \\ \varepsilon_X(x_p, y_p) = -\dfrac{R}{r_p}\dfrac{\partial W(x_p, y_p)}{\partial x_p} \\ \dfrac{R}{r_p} = \dfrac{-1}{n' u'} \approx 2f/\#_W, \end{cases} \tag{2.8}$$

where n' and u' are the image space index and marginal ray angle respectively, R is the image distance (radius of reference sphere), r_p is the physical pupil radius, $\dfrac{\partial W(x_p, y_p)}{\partial y_p}$ and $\dfrac{\partial W(x_p, y_p)}{\partial x_p}$ are the wavefront's slope in x and y [28].

The RMS spot size is a common metric for the quality of an image point. The RMS spot size is the square root of the sum of the squares of each ray's distance (in the transverse ray error) from the center of gravity of rays, divided by the total number of rays. The spot centroid relative to the reference image location is determined by averaging the ray errors.

$$\bar{\varepsilon}_Y = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{Yi} \quad \bar{\varepsilon}_X = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{Xi}. \tag{2.9}$$

A good measure of spot size is the RMS spot size where the rays' locations relative to the spot centroid are summed over the pupil [28].

$$\begin{aligned} RMS_Y &= \left[\frac{1}{N}\sum_{i=1}^{N}(\varepsilon_{Yi} - \bar{\varepsilon}_Y)\right]^{1/2} \\ RMS_X &= \left[\frac{1}{N}\sum_{i=1}^{N}(\varepsilon_{Xi} - \bar{\varepsilon}_X)\right]^{1/2}. \end{aligned} \tag{2.10}$$

And the radial RMS spot size is

$$RMS_R^2 = RMS_X^2 + RMS_Y^2 \tag{2.11}$$

2.6.4 Structure function and connection to optical performance

Atmospheric turbulence is usually characterized by structure function. The errors on telescope's optics should be smaller than the atmospheric effects. Analyzing optical surfaces with SF provides means of comparing optical surfaces to atmospheric turbulence at all spatial length scales [55].

Structure function is also a complementary method for characterizing and representing the spatial content of the optical surface. Unlike PSD, SF does not require the pre-processing steps of windowing and de-trending. Area SF provides a better visualization of the spatial content of the surface as a function of τ (separation) compared to ACF [60]. For statistically stationary surfaces, SF can be connected to optical performance since SF is connected to PSD for such surfaces (Eqn. 2.4) and length scale measures can be obtained [64]. Stationary surfaces are constant in their statistical parameters over space.

2.6.5 ACF and connection to optical performance

For smooth and statistically stationary surfaces, ACF contains similar information as SF but since ACF depends on the mean-plane (piston), it is not a suitable representation for non-stationary surfaces where the statistical characteristics of any sub-aperture are position dependent [57]. For stationary surfaces that have the same sub-aperture statistical characteristics, the ACF is related to both SF and PSD (Eqn. 2.5). For stationary surfaces, a PSD can be obtained from ACF that is identical to the PSD from the profile.

The Fourier transform of the autocovariance function provides the angular properties of scattered light from the surface [98]. The point spread function that is the image of a point source through the aperture, is the inverse Fourier transform of the optical transfer function. The optical transfer function is the autocorrelation of the aperture function that

contains the optical surface error information [25, 99]. The ACF is an important tool to characterize rough surfaces [63]. It is especially a useful representation for spatial variation in surfaces.

2.7 Intuitiveness

This feature of the mathematical analysis is very important. If the analysis is intuitive, the results can be easily interpreted by fabrication engineers and shop floor operators without requiring extensive training. Also, they will help the fabricator decide how to converge to specification by relating the characterization results to fabrication parameters such as tool radius and feed rate in diamond turning process, the tool size, and overlap in the programmed path of the polishing tool or size of the removal footprint in the MRF machine. Intuitiveness of the characterization method will make it more feasible to correct the fabrication process based on the analysis of the measured surface. If the analysis is intuitive, information about the length scale, magnitude, and type of the MSF texture will be easily obtained and extracted from the analysis.

2.7.1 Intuitiveness of PSD

PSD is the most common analysis for MSF texture due to its intuitiveness. PSD is intuitive in separating different spatial bandwidths of the surface and extracting the RMS of the surface through different spatial frequencies. It is also relevant to the surface scattering behavior and its optical performance [11, 96].

2.7.2 Intuitiveness of Zernike polynomials analysis

Zernike analysis is not currently intuitive for analysis of MSF texture without prior training for the shop floor engineers. However, Zernike analysis is readily intuitive for analysis of figure (form) errors and easily interpreted by fabricators. Additionally, the fit

coefficients of these polynomials can be added in quadrature to present the overall RMS of the surface. This condition is true if the polynomials are RMS normalized (orthonormal). C. J. Evans *et al*. have shown an intuitive way of representing the surface aberration in terms of Zernike polynomials. They have shown a part's Zernike coefficients according to their radial and azimuthal orders. Then the non-rotationally symmetric terms are added in quadrature to give the amplitude of the aberration independently of its angular orientation. The same grouping is performed on azimuthal orders to provide angular frequency [88]. In the next chapter, we will show results of Zernike analysis on synthesized data in order to build some intuition to interpret Zernike analysis for MSF textures. We will show how it is possible to intuitively extract information about orientation, length scale and amplitude of the MSF texture.

2.7.3 Intuitiveness of Forbes polynomial analysis

Similar to Zernike analysis, Forbes polynomials for MSF analysis are not readily intuitive for fabricators without training on how to extract the relative information from the Q-spectra. Forbes has analyzed synthetic structures with the spectral method fitting to Q-polynomials. The diagonal banding in the Q-spectra indicates a raster-like pattern. The highest ridge of the distribution happens at $t=\pi C$ where C is the number of cycles across aperture [49, 100]. The Q-spectra can be filtered in terms of m and t for separation of angular and raster frequency respectively [49].

2.7.4 Intuitiveness of structure function

Since the SF is a direct statistical measure of the average height difference on the surface as a function of separation, its interpretation is more intuitive compared to the ACF.

However, for analysis of MSF texture, the structure function is less intuitive than PSD and polynomial analysis. The connection between SF and RMS gradient has been made in [64].

$$SF(r \to 0) = \frac{r^2}{2}(\nabla z)^2_{RMS}, \tag{2.12}$$

Where $r = |P_1 - P_2|$ and p_1 and p_2 are two different points on the surface. SF provides a better visualization of the spatial content of the surface as a function of separation and directionality compared to ACF. For stationary surfaces, the ASF and ACF are closely related. For non-stationary surfaces, there is no simple relationship between ASF and ACF. Therefore, a comparison of ACF and ASF can be used to differentiate a stationary and non-stationary surface.

2.7.5 Intuitiveness of ACF

An intuitive and important parameter of ACF is the correlation length. The correlation length is the shift distance at which the ACF is equal to a fixed value typically 1/e (natural logarithm), 0.1 or zero [62]. In periodic structures, ACF reflects the same period as the input surface profile. Area ACF is the sum of terms each of which is the product of two amplitudes and as R. S. Sayles *et al*. pointed out is difficult to understand and not intuitive [57]. Area ACF is strongly dependent on the mean plane and computationally more intensive compared to area SF.

CHAPTER 3 : ZERNIKE POLYNOMIAL REPRESENTATION FOR SPATIAL FREQUENCY CONTENT OF OPTICAL SURFACES

In this chapter, we investigate a Zernike polynomial representation for quantifying the mid-spatial frequency content of the surface. The chapter is organized into several sections. Section 3.1 is complementary to 1.3.2 and introduces Zernike polynomials in more detail. Section 3.2 includes a discussion on the recursive formulas for stable generation of high order Zernike polynomials. In section 3.3.1, we show the results of least square fitting to synthetic maps to build intuition on how to interpret coefficient plots. Zernike fits to data from measured surfaces with certain characteristics are shown in section 3.3.2. We will show the filtering capability of Zernike polynomials and present examples of filtered bands based on symmetry properties of Zernike polynomials. In 3.4, we will look at the phase between different filtered bands and their effect on the overall height map. In 3.5, there will be a discussion about the convergence of RMS fit residual to investigate the effectiveness of a Zernike polynomial representation for quantifying MSF and we will discuss limits of least square fitting.

3.1 Zernike polynomials

Any wavefront within a circular aperture can be described as a weighted sum of Zernike polynomials such as [101]

$$W = \sum_{n_z} \sum_{m} \sigma_{n_z}^{\pm m} Z_{n_z}^{\pm m}(\rho, \theta), \tag{3.1}$$

and

$$Z_{n_z}^{m} = \begin{Bmatrix} Z_{n_z}^{m}(\rho)\cos(m\theta) & m \geq 0 \\ Z_{n_z}^{m}(\rho)\sin(m\theta) & m < 0 \end{Bmatrix} \tag{3.2}$$

where n_z are nonnegative integers and ($(n_z - |m|) \geq 0$ and even). In order to avoid confusion with Forbes polynomials, the Zernike radial orders are indicated with the parameter n_z. The parameter *m* is related to the azimuthal frequency and the index n_z is the radial degree or the order of the polynomial because it indicates the highest power of ρ in the polynomial [102]. ρ is the radial distance ($0 \leq \rho \leq 1$), and φ is the azimuthal angle between 0 and 2π. The radial polynomials $R_{n_z}^{m}(\rho)$ are defined as

$$R_{n_z}^{m}(\rho) = N_{n_z}^{m} \sum_{k=0}^{(n_z-m)/2} \frac{(-1)^k (n_z - k)!}{k!(\frac{n_z+m}{2}-k)!(\frac{n_z-m}{2}-k)!} \rho^{n_z-2k}. \tag{3.3}$$

The normalization constant $N_{n_z}^{m}$ is defined as

$$N_{n_z}^{m} = \begin{Bmatrix} 1 & non-normalized \\ \left(\frac{2(n_z+1)}{1+\delta_{m0}}\right)^{1/2} & RMS-normalized \end{Bmatrix}, \tag{3.4}$$

where

$$\delta_{m0} = \begin{cases} 0, \text{ if } m \neq 0 \\ 1, \text{ if } m = 0 \end{cases}. \tag{3.5}$$

The normalization constant makes the radial polynomials $R_{n_z}^{m}(\rho)$ normalized such that $R_{n_z}^{\pm m}(1) = 1$ for all values of n_z and m. In RMS normalization, the polynomials are orthonormal, and the sum in quadrature of the fit coefficients is equal to the RMS of the surface. Expressed differently, the RMS of the fitted surface after subtraction of a specific

fitted term is given by the difference in quadrature of the fitted surface and the specific term. An orthonormal set of functions $\phi_0(x),\ \phi_1(x),\ ...,\ \phi_l(x)$ is defined by the relation

$$\langle \phi_n,\ \phi_m \rangle = \int_a^b \phi_n(x)\phi_m(x)\ \mathrm{d}\alpha(\mathrm{x}) = \delta_{nm}, \qquad n,\ m = 0,\ 1,\ 2,\ ...,\ l. \tag{3.6}$$

Here, the functions $\phi_n(x)$ are real valued and linearly independent [103].

There are different indexing schemes for Zernike polynomials, single indexing and double indexing [104]. The most common double indexing schemes are given by Born and Wolf (sometimes referred to as standard ordering) and by Loomis (referred to as Fringe ordering) [47, 48]. In this dissertation, we use the standard Born and Wolf double indexing scheme unless stated otherwise. Describing a surface using the RMS normalized Zernike polynomials, has the advantage that each coefficient represents the contribution of that specific polynomial. Table 3.1 shows the Zernike polynomial sequence, orders one through seven using standard ordering. Please note that the azimuthal order m in Table 3.1 is indicative of $\sin(|m|\theta)$ and the RMS column is indicative of the normalization factor in Eqn. 3.4.

Table 3.1 Zernike polynomials in Polar coordinates (the standard set) [45]

n_z	m	Sets	RMS	Orthogonal polynomials	Aberration names
0	0	Z_0^0	$\sqrt{2}$	1	piston
1	-1	Z_1^{-1}	2	$\rho\sin\theta$	tip
1	1	Z_1^1	2	$\rho\cos\theta$	tilt
2	-2	Z_2^{-2}	$\sqrt{6}$	$\rho^2\sin 2\theta$	astigmatism
2	0	Z_2^0	$\sqrt{3}$	$2\rho^2-1$	Power (defocus)
2	2	Z_2^2	$\sqrt{6}$	$\rho^2\cos 2\theta$	astigmatism
3	-3	Z_3^{-3}	$2\sqrt{2}$	$\rho^3\sin 3\theta$	trefoil
3	-1	Z_3^{-1}	$2\sqrt{2}$	$(3\rho^3-2\rho)\sin\theta$	coma
3	1	Z_3^1	$2\sqrt{2}$	$(3\rho^3-2\rho)\cos\theta$	coma
3	3	Z_3^3	$2\sqrt{2}$	$\rho^3\cos 3\theta$	trefoil
4	-4	Z_4^{-4}	$\sqrt{10}$	$\rho^4\sin 4\theta$	
4	-2	Z_4^{-2}	$\sqrt{10}$	$(4\rho^4-3\rho^2)\sin 2\theta$	
4	0	Z_4^0	$\sqrt{5}$	$6\rho^4-6\rho^2+1$	
4	2	Z_4^2	$\sqrt{10}$	$(4\rho^4-3\rho^2)\cos 2\theta$	
4	4	Z_4^4	$\sqrt{10}$	$\rho^4\cos 4\theta$	
5	-5	Z_5^{-5}	$2\sqrt{3}$	$\rho^5\sin 5\theta$	
5	-3	Z_5^{-3}	$2\sqrt{3}$	$(5\rho^5-4\rho^3)\sin 3\theta$	
5	-1	Z_5^{-1}	$2\sqrt{3}$	$(10\rho^5-12\rho^3+3\rho)\sin\theta$	
5	1	Z_5^1	$2\sqrt{3}$	$(10\rho^5-12\rho^3+3\rho)\cos\theta$	
5	3	Z_5^3	$2\sqrt{3}$	$(5\rho^5-4\rho^3)\cos 3\theta$	
5	5	Z_5^5	$2\sqrt{3}$	$\rho^5\cos 5\theta$	
6	-6	Z_6^{-6}	$\sqrt{14}$	$\rho^6\sin 6\theta$	
6	-4	Z_6^{-4}	$\sqrt{14}$	$(6\rho^6-5\rho^4)\sin 4\theta$	
6	-2	Z_6^{-2}	$\sqrt{14}$	$(15\rho^6-20\rho^4+6\rho^2)\sin 2\theta$	
6	0	Z_6^0	$\sqrt{7}$	$20\rho^6-30\rho^4+12\rho^2-1$	
6	2	Z_6^2	$\sqrt{14}$	$(15\rho^6-20\rho^4+6\rho^2)\cos 2\theta$	
6	4	Z_6^4	$\sqrt{14}$	$(6\rho^6-5\rho^4)\cos 4\theta$	
6	6	Z_6^6	$\sqrt{14}$	$\rho^6\cos 6\theta$	
7	-7	Z_7^{-7}	4	$\rho^7\sin 7\theta$	
7	-5	Z_7^{-5}	4	$(7\rho^7-6\rho^5)\sin 5\theta$	
7	-3	Z_7^{-3}	4	$(2\rho^7-30\rho^5+10\rho^3)\sin 3\theta$	
7	-1	Z_7^{-1}	4	$(35\rho^7-60\rho^5+30\rho^3-4\rho)\sin\theta$	
7	1	Z_7^1	4	$(35\rho^7-60\rho^5+30\rho^3-4\rho)\cos\theta$	
7	3	Z_7^3	4	$(2\rho^7-30\rho^5+10\rho^3)\cos 3\theta$	
7	5	Z_7^5	4	$(7\rho^7-6\rho^5)\cos 5\theta$	
7	7	Z_7^7	4	$\rho^7\cos 7\theta$	

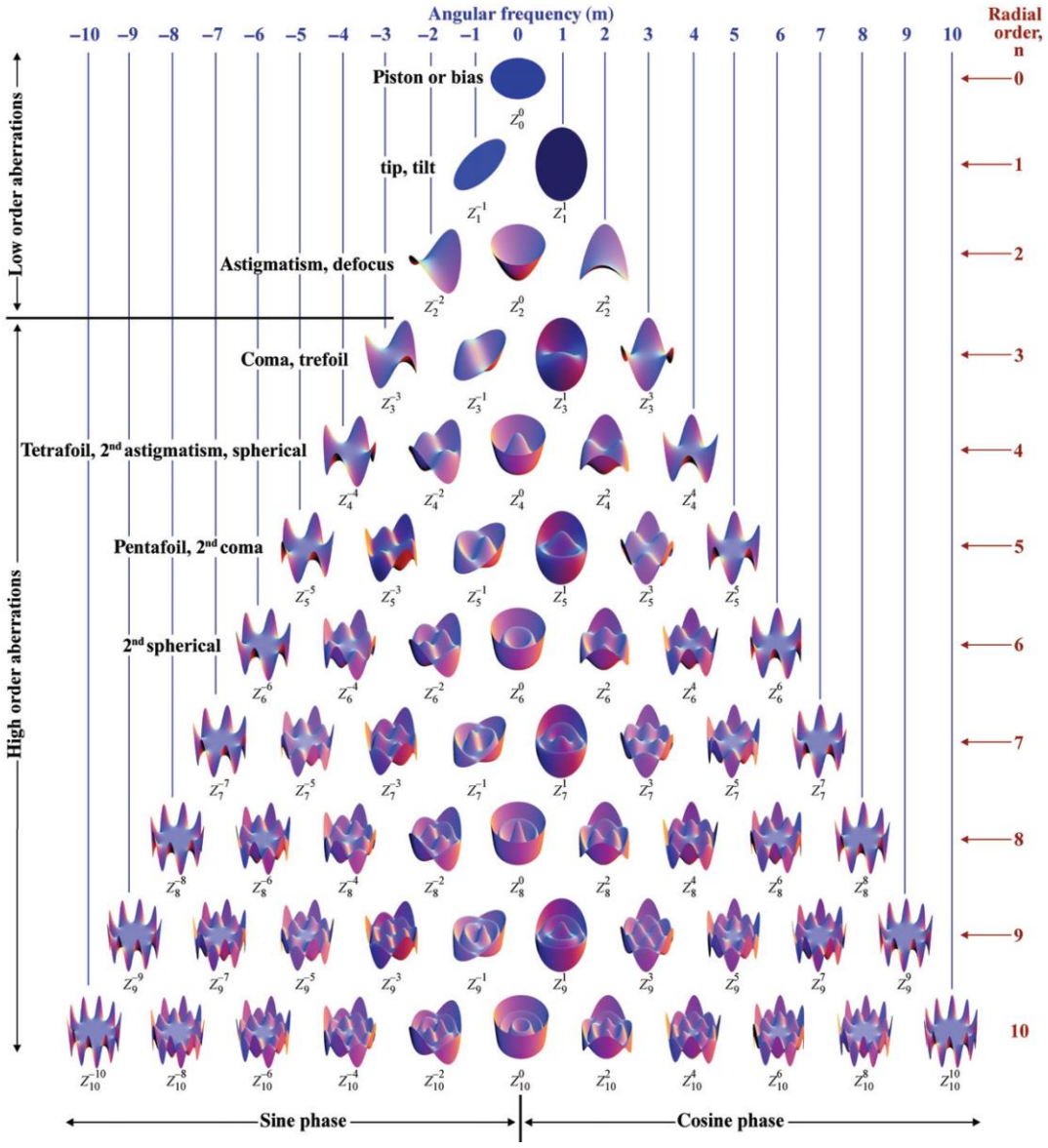


Figure 3.1 Zernike polynomials (standard set) (n≤7, m≤7) [45].

3.2 Recursive generation of high order Zernike polynomials

The radial part of Zernike polynomials is described mathematically by explicit formulas as in Eqn. 3.3. However, these explicit formulas will result in numerical instability for higher orders [87]. Even with using double precision calculations, the results become extremely large especially for values of $\rho \to 1$[87]. This is due to the fact that for $\rho \approx 1$ the significant digits get lost due to cancellation in the explicit method. There are about 16

significant digits in a double precision floating point number that occupies 64 bits in computing memory. On the other hand, when using recursive formulas, the rounding error introduced in the nth step only contributes an error $\varepsilon_n p_n$ to the final result which does not increase the errors in the following steps. This is because the error introduced during the calculation of each term follows the same recurrence relation as the polynomials itself [105]. In order to describe a freefrom surface completely through low order out beyond the mid-spatial frequency region, very large orders of polynomials are required. To achieve numerical stability, recurrence relations should be applied to avoid round-off errors. G. W. Forbes and I. Kaya *et al.* have taken this approach to fit to high orders of polynomials [49, 93]. Zernike polynomials are related to Jacobi polynomials as described in Eqn. 3.7 [86]. Jacobi polynomials can be derived using a three-term recurrence relation shown in Eqn. 3.8. The first two terms in the recurrence relation can be found in Eqn. 3.9. The coefficients a_n, b_n, c_n are described in Eqns. 3.10 to 3.12 [106]. Using recursive Jacobi polynomials for generation of robust and large order Zernike polynomials was first introduced by G. Forbes [87].

$$R_{n_z}^{m}(\rho) = \rho^m P_{\frac{(n_z-m)}{2}}^{(0,m)}(2\rho^2 - 1) \tag{3.7}$$

$$P_{n+1}(x) = (a_n + b_n x)P_n(x) - c_n P_{n-1}(x) \tag{3.8}$$

$$P_0^{0,m}(x) = 1, P_1^{0,m}(x) = \tfrac{1}{2}(m+2)x - \tfrac{1}{2}m \tag{3.9}$$

$$a_n = \frac{-m^2(2n+m+1)}{2(n+1)(n+m+1)(2n+m)} \tag{3.10}$$

$$b_n = \frac{(2n+m+1)(2n+m+2)}{2(n+1)(n+m+1)} \tag{3.11}$$

$$c_n = \frac{n(n+m)(2n+m+2)}{(n+1)(n+m+1)(2n+m)} \tag{3.12}$$

The radial order n (shown as n_z) in Zernike polynomials is related to the radial order in Forbes' representations, n_f through the relation $n_f = \frac{n_z - m}{2}$ [107]. The consequences of not using recurrence relations for high-order Zernike polynomials are illustrated in Figure 3.2 of reference [87] and Figure 2 of reference [107]. Figure 3.2 shows plots of high order rotationally invariant Zernike polynomial $Z_{25}^0(x)$ in the range of $0.95 < x < 1$ [87].

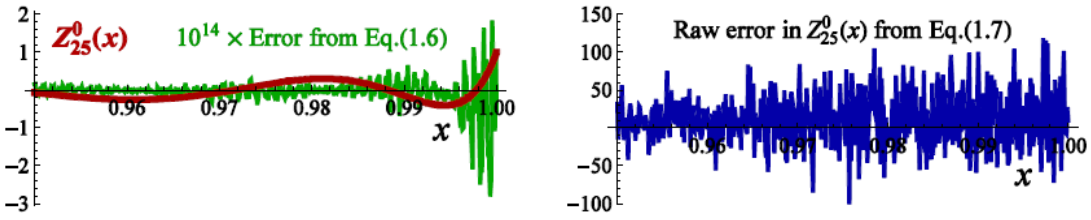


Figure 3.2 Plots of high order rotationally symmetric Zernike polynomial $Z_{25}^0(x)$ in the range of $0.95 < x < 1$. The red curve on the left shows the correct plot of $Z_{25}^0(x)$ and the green curve is 10^{14} times the errors produced by recursive formulas. The blue curve on the right shows the catastrophic error when using explicit equations [87].

Figure 3.3 (a) is an example of a high-order Zernike polynomial generated using the recursion formulas. In Figure 3.3 (b) we show the difference in the peak to valley (PV) amplitude of Zernike maps generated using explicit formulas *vs.* recursive formulas such that $h_{n_z}^m = \frac{PV\left[(Z_{n_z}^m)_{explicit} - (Z_{n_z}^m)_{recursive}\right]}{PV\left[(Z_{n_z}^m)_{recursive}\right]}$. This map represents the fractional error in the PV normalized by the recursive value. It also represents the extent to which the explicit method is valid. The explicit formulas start to become unstable from $n_z \approx 40$. The areas where the maximum amplitude of explicitly generated polynomials are more than 10 times larger than

the maximum amplitude of recursively generated polynomials (the correct values) are set to the same dark red color to show a clear boundary where the explicit generation method is invalid.

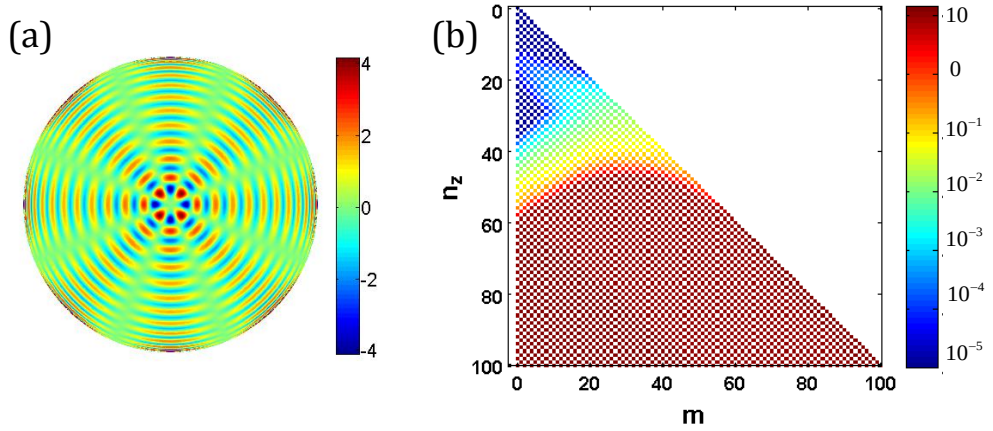


Figure 3.3 (a) Illustration of high- order Zernike polynomial $Z_{50}^{4}(\rho,\varphi)$ using the recursive generation approach. (b) Each data point in this map is the difference in the PV amplitude between the explicit and recursive generated Zernike polynomials of radial order n_z and azimuthal order m.

We have used least-square fitting methods to calculate the fit coefficients of Zernike polynomials. Fitting to large order Zernike polynomials here is carried out in a sequential manner to minimize error. i.e., each polynomial term multiplied by its fit coefficient is subtracted from the map before performing the next fit. The results of mid-spatial frequency characterization using Zernike polynomials are presented in section 3.3.

3.3 Mid-spatial frequency representation using Zernike polynomials

Using large order Q-polynomials for capturing mid-spatial frequency was first reported by G. Forbes [49, 84]. In this section, pure sinusoidal patterns are fit to high orders of Zernike polynomials. These synthetic maps resemble mid-spatial frequency patterns left as a result of different freeform fabrication processes. Raster, ring-like and spokes patterns

can be generated as a result of different manufacturing processes. The intent of our analysis is to find out how Zernike polynomials describe a different aspect of MSF structure such as type (raster, ring-like, etc.), orientation, length scale of the period, magnitude of different length scales and their relative phase. When possible, we will also compare Zernike analysis of MSF with Forbes Q-polynomials.

As mentioned earlier, surface errors within circular apertures can be described by Eqn. 3.1 that can also be written as

$$W = \sum_{n_z} \sum_{m} R_{n_z}^{m}(\rho)\left[\alpha_{n_z}^{m} \cos(m\varphi) + \beta_{n_z}^{m} \sin(m\varphi)\right]. \tag{3.13}$$

The coefficients $\alpha_{n_z}^{m}$ and $\beta_{n_z}^{m}$ are calculated for high orders of Zernike polynomials for a range of signature specific surfaces. The result of the fit can be displayed with a coefficient map. The root sum square (RSS) of the corresponding α and β coefficients are calculated. $\sqrt{\left(\alpha_{n_z}^{m}\right)^2 + \left(\beta_{n_z}^{m}\right)^2}$ of different orders are plotted using Forbes indexing scheme, i.e. the coefficients are shown for n_f and m [49]. $\alpha_{n_z}^{m}$ is related to cos terms ($m \geq 0$) and $\beta_{n_z}^{m}$ is related to sine terms ($m < 0$). The Cartesian order associated with these coefficients is described by [51, 84]:.

$$t = \begin{cases} 2n + 4, & m = 0, \\ 2n + m, & m > 0. \end{cases} \tag{3.14}$$

The Cartesian order t is indicative of degrees of freedom for describing a freeform surface using polynomials of those order. The RSS of α and β coefficients is the same as $\sigma_{n_z}^{m}$ in Eqn. 3.1 and it shows the amount of each Zernike term in the wavefront for both ($\pm m$) terms. All $\sigma_{n_z}^{m}$ can be added in quadrature to provide the total RMS of the surface.

Fit coefficients of synthetic and real maps are calculated for 300 orders of Zernike polynomials (45451polynomial terms). Fit coefficients that are smaller than 0.001 nanometers are not displayed in the maps.

3.3.1. Synthetic data, pure periodic maps

The first synthetic map we discuss is a vertical linear sinusoidal pattern with 25 cycles across the aperture with the height map $z = \sin(\frac{25\pi x}{\rho_{\max}} + \delta)$ where the initial phase δ is $\pi/4$. For vertical or horizontal sinusoidal patterns, the values of n_f and m where the coefficients are largest happen at unique values corresponding to the length-scale of the periodicity. For 25 cycles over the aperture, the peak value corresponds to Cartesian order $t \approx 75$. This is along a diagonal in the n_f and m map as shown in Figure 3.4. G. Forbes calculated the Q-spectra of the same sinusoidal map to compare the efficiency of Q-polynomials for describing MSF to Fourier methods [49]. He observed that the highest ridge of Q-spectra occurs near $t \approx \pi C$ where C is the number of cycles across the aperture [49]. The Zernike polynomial representation shows the same peak characteristics at the same Cartesian order. This is not surprising, given the similarities between these two polynomial descriptions. The PV of the height error in Figure 3.4 (a) is 2 μm with the RMS of 0.71 μm. We have calculated the root sum square (RSS) of the coefficients for different values of t diagonally. Figure 3.4 (c) shows that the RSS of the fit coefficients peaks at a thick band around $t \approx 75$. It is interesting to notice that the RSS plot shows that the RSS drastically drops after the peak but not before the peak at $t \approx 75$.

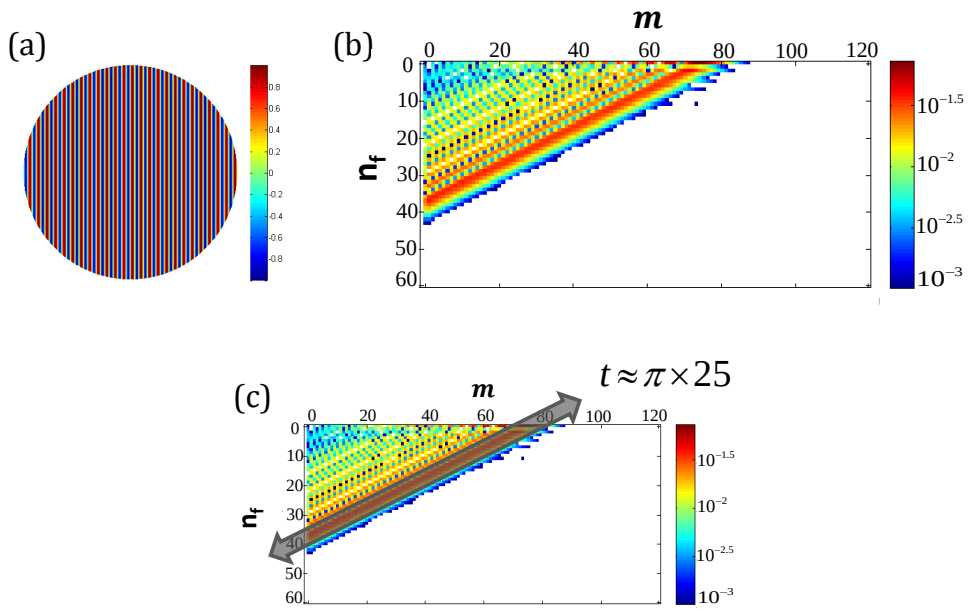


Figure 3.4 (a) Height map of sinusoidal raster pattern with 25 cycles within the aperture. (b) Fit coefficient plot resulted from fitting the map on the left to 120 orders of Zernike polynomials (7381 terms of basis elements). The peak of the coefficients occurs at Cartesian order $t \approx 75$. (c) The highlight of the peak of diagonal band.

To calculate the magnitude of the $C = 25$, we have calculated the RSS along the thick band of $73 < t < 78$ shown in Figure 3.5 (a). The amplitude of this band is 0.47 µm which is about 67% of the total RMS of the surface. The thick band in the coefficient map (rather than a sharp band/peak) is a sign that the polynomials are not a particularly natural basis set for a periodic-like structure such as raster-like patterns. For comparison with Q-spectra please refer to Figure 3 in reference [49].

In order to speculate on the effectiveness of this method in providing information on the length scale and amplitude of the structures on the surface, we can compare this analysis to PSD that is the most common approach in retrieving the surface finish parameters. Looking at the PSD of a linear trace across Figure 3.4 (a), we see a peak at $f_x = 12.5 \frac{1}{mm}$ that is equal to the spatial period of $\Lambda = 0.08mm$. The area under the PSD curve is equal

to the square of the RMS of the surface (σ^2) and can be computed numerically [62]. An alternative approach to finding RMS of the surface at a specific frequency band is applying a band-pass Fourier filter to the surface and calculating the RMS of the results.

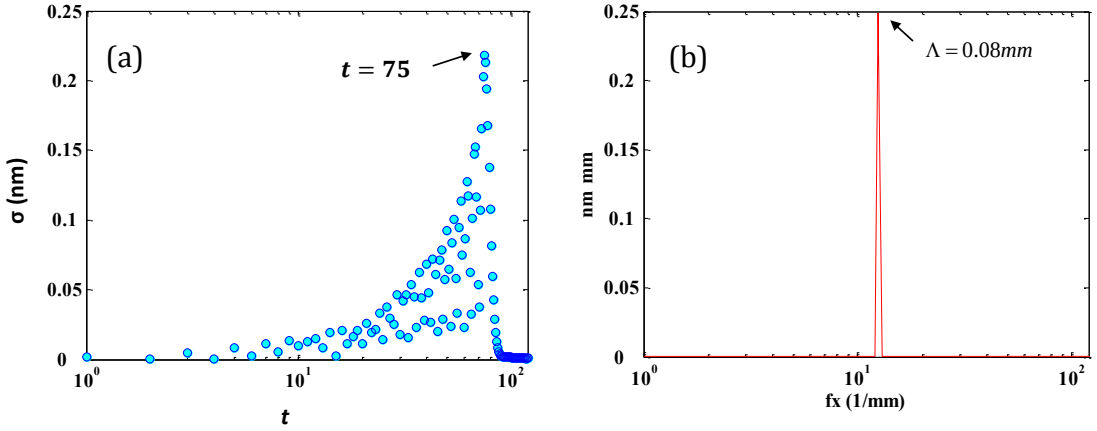


Figure 3.5 (a) RSS of diagonal bands along Cartesian order t of the coefficient map of sinusoidal raster pattern with 25 cycles across the aperture. (b) Linear PSD of the same sinusoidal raster pattern.

The coefficients of Zernike polynomials are also calculated for a radial sinusoidal pattern $z = \sin(\frac{2\pi C\rho}{\rho_{\max}})$, with $C = 26$. The peak of the coefficient distribution occurs at rotationally invariant orders where the azimuthal order m is zero. For capturing and describing ring-like structures, superposition of rotationally invariant Zernike polynomials is needed, hence several terms in the $m = 0$ band. Plotting the fit coefficients at $m = 0$ in Figure 3.6 (b), shows that the peak of the coefficients in ring-like structures is weakly related to the number of cycles across the aperture. In this case, the peak occurs at $n_f \approx 38$ or $t \approx 80$ that is related to the number of rings across the aperture through $C \approx \frac{t}{\pi} = \frac{2n_f + 4}{\pi} \approx 26$ [49] and is a special case of Figure 3.4. In this case, the peak is not

a sharp one and the values at $n_f = 34$ and $n_f = 37$ are very close to $n_f = 38$. As expected, calculating the RSS of the coefficients for rotationally invariant terms i.e. along $m = 0$ provides a value of 0.71 μm that is the same as the directly calculating the RMS of the surface with a PV of 2 μm.

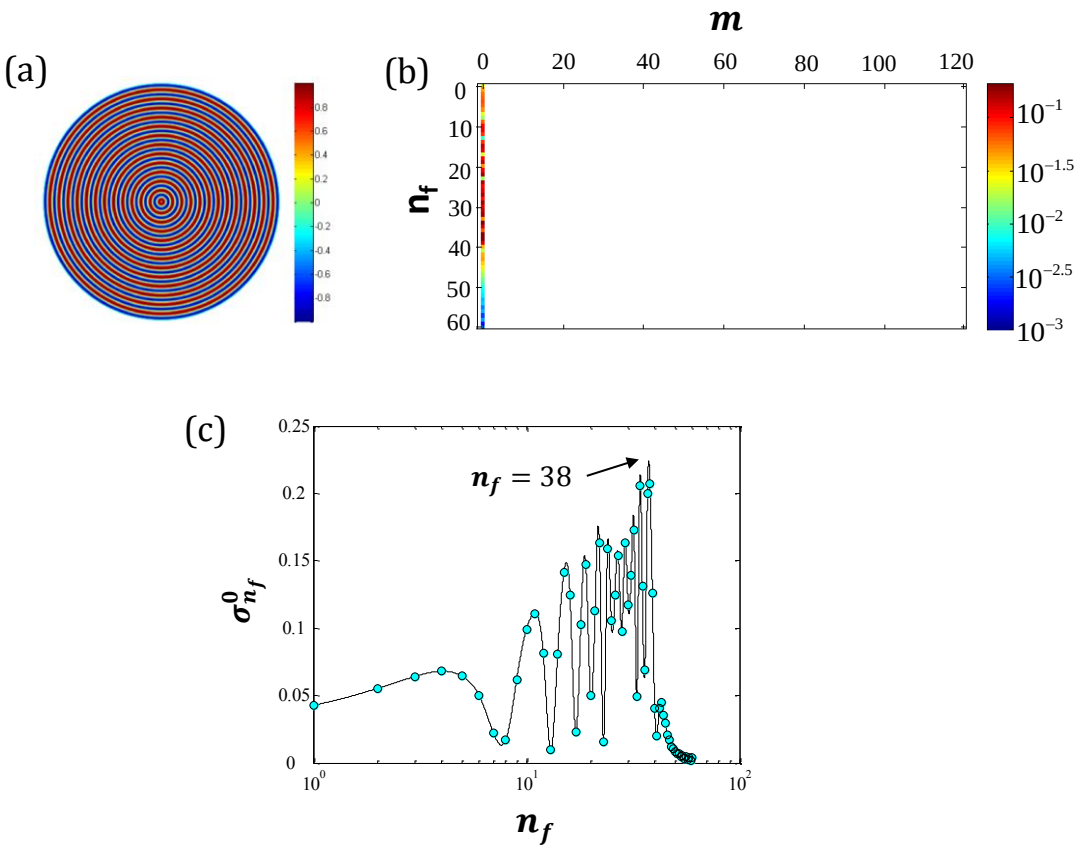


Figure 3.6 (a) Radial sinusoidal pattern with 26 cycles across the aperture. (b) The peak of the coefficients happens at rotationally invariant terms or $m = 0$. (c) Coefficients along n_f show a peak at $n_f \approx 38$ that corresponds to 26 cycles across the aperture. RSS along $m = 0$ gives a value of 0.71 μm that is the same as RMS of the sinusoidal pattern calculated directly.

For a spokes pattern, the peak of the distribution is at an azimuthal order m, where m equals the number of azimuthal periods in the map, in this case at $m = 9$. This follows from the definition and symmetry of the Zernike polynomials. The largest coefficient in

the $m=9$ band happens at $n_f=0$ ($n_z=9$). The amplitude of coefficients gradually decreases for higher orders of n_f . RSS of the coefficients along $m=9$ in Figure 3.7 (b) is 0.71 µm that is the same as the RMS of the surface calculated directly.

The results of fitting synthetic maps to high orders of Zernike polynomials provides information on the type, length scale and amplitude of MSF signatures through coefficient maps. Raster, ring-like and spokes patterns have diagonal, rotationally invariant and vertical peaks in the coefficient maps, respectively. The length scale or number of cycles across the aperture can be estimated from the location of the peak of the coefficients at Cartesian order t or the number m for spokes patterns. The amplitude of the dominant features can be derived through RSS of the peak values of the coefficients. Forbes Q-polynomials behave similarly when used to describe MSF error [49]. The Zernike coefficients can be used to estimate the RMS of the height map whereas the Q-spectra (coefficients of Q-polynomials) give access to RMS gradient [50].

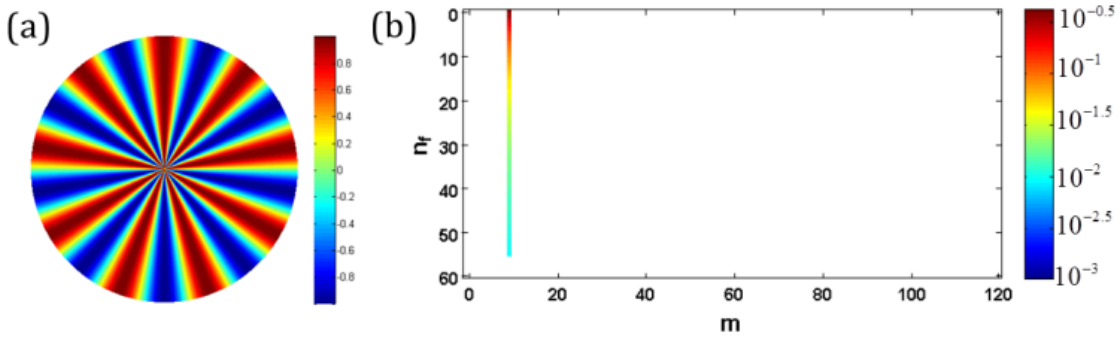


Figure 3.7 (a) Spokes pattern with 9 cycles across 2π azimuthal angle. (b) A coefficient peak occurs at $m=9$.

In Figure 3.8, the Q-spectra associated with some synthetic sinusoidal maps are plotted. The Q-spectra of the vertical sinusoidal map with 25 cycles across aperture -Figure 3.8 (a)- is very similar to the Zernike coefficient plot of the same surface map. You can see a thick band of large values and the location of the dark peak is at the same Cartesian orders. For the radial patterns, the peaks happen at $m = 0$ and that's the same case for Zernike polynomials. For spokes patterns similar to Figure 3.7 (a), due to the harsh singularity of the error map at the origin, instead, as suggested through personal communication with G. W. Forbes, the Q-spectra of the function $\phi = \rho^9 \cos(9\theta)$ is calculated. The Q-spectra of these maps were calculated using an open-source Python package that implements the algorithm described in reference [49] [49, 108]. The Zernike coefficients can be used to estimate the RMS of the height map whereas the Q-spectra (coefficients of Q-polynomials) give access to RMS gradient [49, 100].

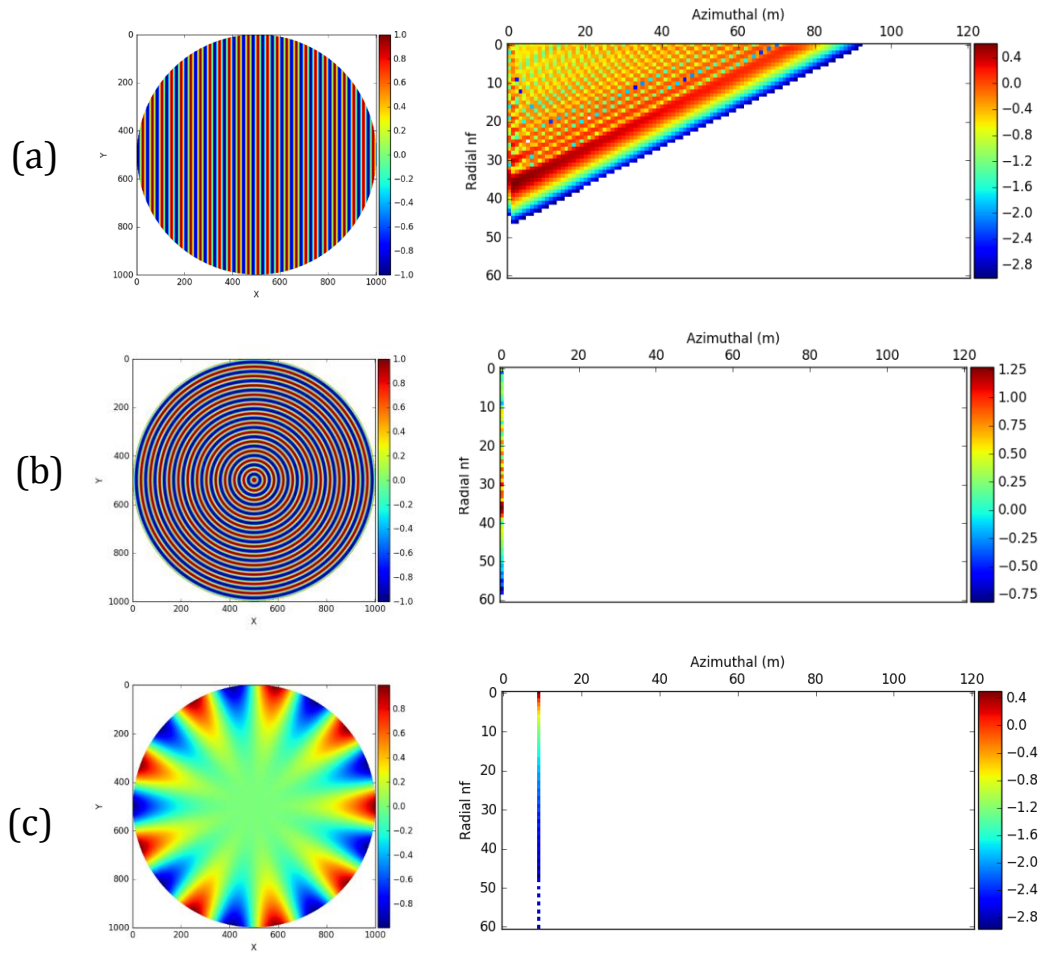


Figure 3.8 Q-spectra of (a) vertical sinusoidal map with 25 cycles across aperture with an initial phase of $\pi/4$. (b) Radial sinusoidal map with 26 cycles across aperture. (c) Spokes pattern with 9 spokes.

3.3.2. Measured surfaces

In this section, there are examples of fit coefficient amplitude plots of measured surfaces with different MSF texture and orientation. These surface errors have been selected from the pool of in-house manufactured samples at UNC Charlotte and samples provided by industry affiliates due to their representative MSF characteristics. For example, the data in Figure 3.9 (a) shows MSF errors that are dominated by azimuthal and spoke-like texture, whereas in Figure 3.11 (a) and Figure 3.13 (a), the MSF errors have radial and linear texture, respectively. Units are in nanometers in all real surface maps.

Figure 3.9 (a) shows the error map of an aluminum diamond turned part and Figure 3.9 (b) shows its coefficient map. It is interesting to observe that there are several azimuthal frequency peaks in this map. Figure 3.9 (c) shows the RSS of the coefficients along each azimuthal frequency where we can clearly read the location and RMS amplitude of each azimuthal frequency. The dominant azimuthal frequencies in this map are at *m=0, m=14, m=15, m=22,* and *m=24* with RMS amplitudes of 31.6 *nm,* 9.1 *nm,* 4.1 *nm,* 3.1 *nm* and 7.4 *nm*, respectively. Although extracting information about length scale and amplitude of pure sinusoidal structures is straightforward from the PSD representation, it is not straightforward for real surfaces like the one in Figure 3.9 (a). Aside from the pre-processing steps and loss of information on the low order form errors due to detrending, extracting information about length scale and amplitude of mid-spatial frequencies involve slicing through the two-dimensional PSD or integrating the two-dimensional PSD into a one-dimensional PSD [41, 42]. In order to do so, prior insight into the direction of MSF texture is required. For example, selecting a horizontal slice of two-dimensional PSD is suitable for linear MSF texture but will not provide information on the spokes patterns present in the surface.

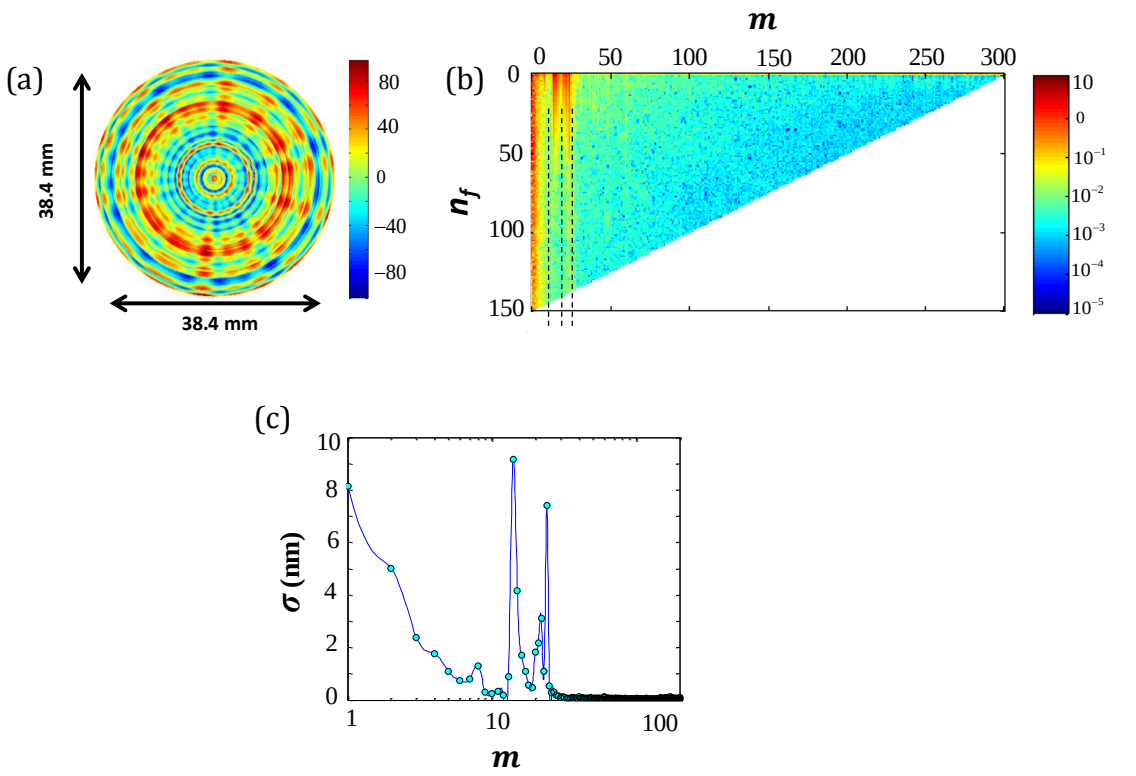


Figure 3.9 (a) Error map of aluminum diamond turned part with dominant spokes patterns of different frequencies. (b) Zernike fit coefficient map of the surface in (a). (c) RSS Zernike coefficient map revealing the peaks at different azimuthal bands. Units are in nm.

In order to showcase the varying azimuthal frequency content of the error map in figure 3.9 (a), we have selected different azimuthal bands of the structure shown by the dashed line in figure 3.9 (b). The RMS of the filtered surfaces at different bands of $0 \le m \le 8$, $9 \le m \le 15$, $16 \le m \le 27$, and $28 \le m$ are 33.9, 10.3, 9.2 and 5.3 *nm* respectively. Within each band in Figure 3.10, you can see the superposition of close azimuthal frequencies with different phases.

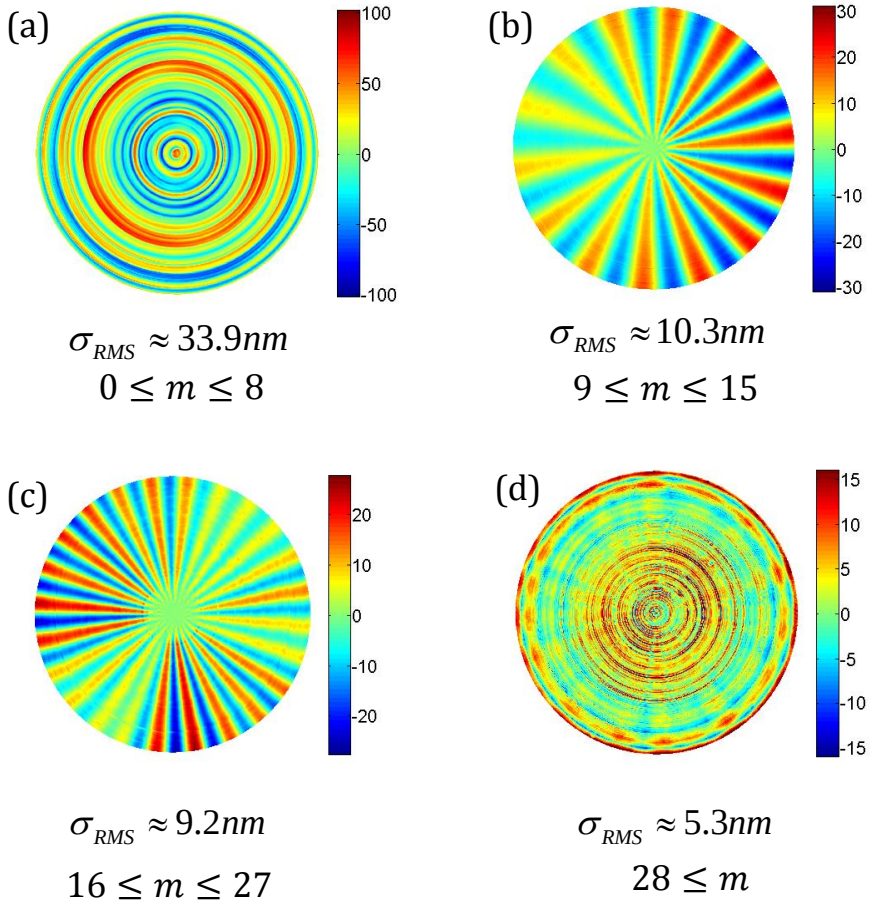


Figure 3.10 (a) Rotationally invariant terms and all the other azimuthal terms between $0 \le m \le 8$ with $\sigma_{RMS} \approx 33.9nm$. (b) Azimuthal orders $9 \le m \le 15$ with $\sigma_{RMS} \approx 10.3nm$. (c) The band containing azimuthal orders between $16 \le m \le 27$ with $\sigma_{RMS} \approx 9.2nm$. (d) The azimuthal bands of $28 \le m$ with $\sigma_{RMS} \approx 5.3nm$.

Figure 3.11 (a) is an error map of a silicon carbide (SiC) spherical mirror, and the corresponding coefficient map is shown to the right. Aside from the dominant presence of large low-order coefficients in the left upper corner due to the form errors, there are rotationally invariant terms that extend to large radial orders of n_f. Looking at the coefficient map, the absence of any linear structure and spokes-like patterns are verified since there are no diagonal or vertical peaks in the coefficient map. This structure is a signature of a turning process and thermal cycles of the process that usually have a larger

length scale. In Figure 3.12 (a) you can see the separation of rotationally invariant terms from the surface. Even after removing the ring-like structure, there is no raster pattern present in the error map. As shown in Figure 3.12, the RMS of the surface is 26.4 nm for $m = 0$ terms. For the non-rotationally invariant content of the surface, the RMS is 24 *nm*.

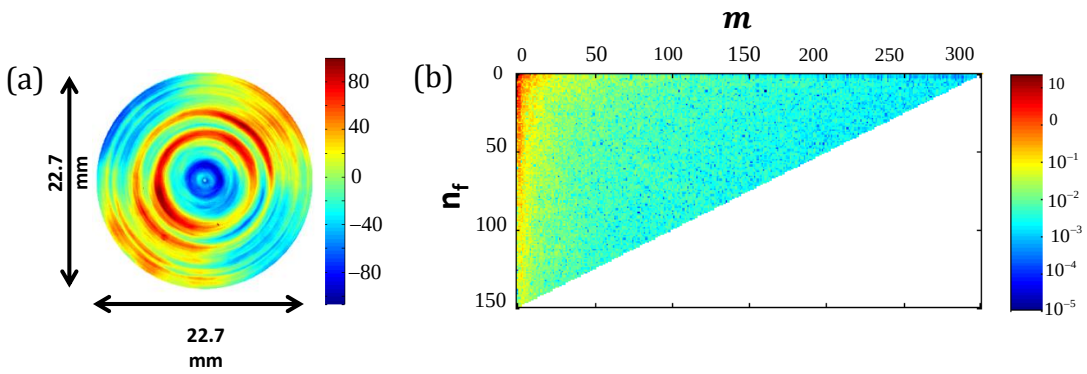


Figure 3.11 (a) The error map of SiC sphere with radial patterns. (b) The RSS Zernike coefficient plot revealing the dominant form errors and rotationally invariant terms.

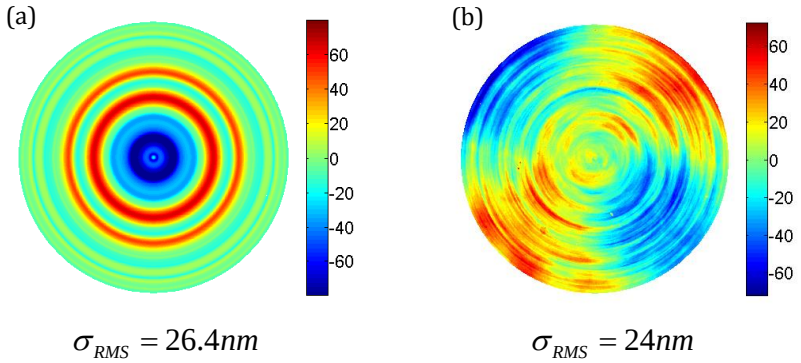


Figure 3.12 (a) Rotationally invariant terms in the height map of SiC sphere with RMS of 26.4 nm. (b) The non-rotationally invariant content of SiC sphere with RMS of 24 nm.

The MSF content in the surface of Figure 3.13 (a) shows signatures of raster-like patterns, and the presence of diagonal peaks in the coefficient plot verifies that. The peaks for this map occur at certain Cartesian orders such as $t \approx 29, 44, 59, 67, 112$. These peaks correspond to approximately 9, 14, 19, 21 and 36 cycles across the aperture that can be

related to machining process parameters such as step-over in the programmed tool path and the geometry of the tool tip in the milling machine. The astigmatism coefficient $\sigma_2^{\pm 2}$ is set to zero for better visibility of higher-order coefficients in Figure 3.13(b). G. Forbes has calculated the Q-spectrum of the same surface and also observed diagonal peaks in the coefficient map [100]. We see the same characteristic in the Zernike spectrum. G. Forbes has isolated different length scales of this data by filtering different diagonal bands of along Cartesian orders of t [100]. Using a bi-linear interpolation has enabled fitting to 250 orders of Q-polynomials [100]. We have isolated different length scales of the surface in Figure 3.13 (a) along different Cartesian orders. The RMS of the filtered surfaces at different bands of $t \le 13$, $25 \le t \le 30$, $35 \le t \le 45$, and $105 \le t \le 115$ are 270, 10.9, 7 and 5 nm respectively. The RMS values of isolated maps for the same bands using Q-polynomials are 270, 11, 6.8, and 4.6 nm respectively [100]. These values are very similar and the small differences might be due to the interpolation process during the Q-polynomial fitting or other pre-processing choices. For comparison of filtering this data using Q-polynomials, please see Figure 3 in reference [100].

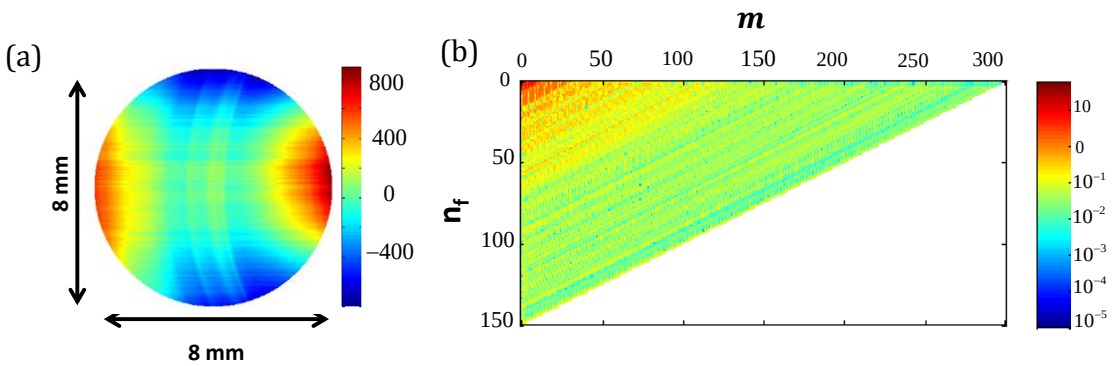


Figure 3.13 (a) Error map of Ge sphere with raster-like MSF structure. (b) The RSS Zernike coefficient map reveals raster patterns at certain Cartesian orders of t peaking diagonally in the coefficient map.

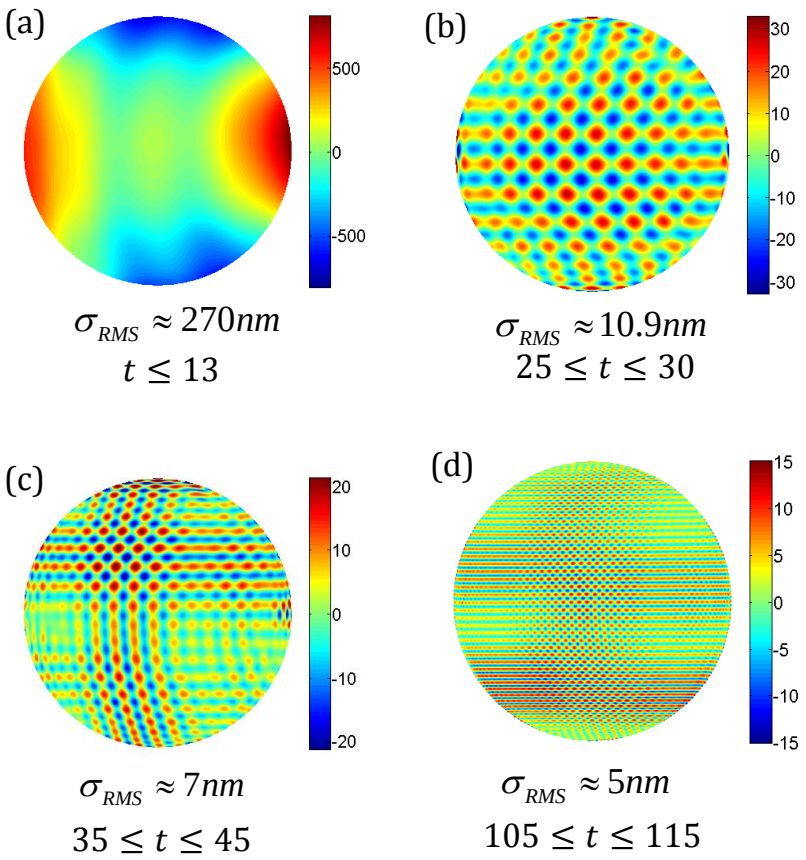


Figure 3.14 (a) Surface maps showcasing different diagonal bands of the germanium sphere in Figure 3.13 (a). (a) Diagonal bands of $t \leq 13$ with $\sigma_{RMS} \approx 270nm$. (b) Cartesian orders of $25 \leq t \leq 30$ with $\sigma_{RMS} \approx 10.9nm$. (c) Cartesian orders between $35 \leq t \leq 45$ with $\sigma_{RMS} \approx 7nm$. (d) Cartesian bands of $105 \leq t \leq 115$ with $\sigma_{RMS} \approx 5nm$.

3.4 Phase, orientation of spatial frequency components

The coefficient maps provide information on the type and amplitude of the MSF texture but do not provide any information on the orientation of the MSF texture. In principle, this information is contained in the relative weighting of the sine and cosine terms in the Zernike representation. That is, if the surface is composed of multiple length scales or frequency components, then the relative phase between the constituent spatial frequencies in the azimuthal θ direction can be extracted through an additional step. Eqn. 3.13 can be rewritten as

$$W = \sum_{n}\sum_{m} R_{n_z}^{m}(\rho) a_{n_z}^{m} \cos(m\theta - \delta_{n_z}^{m}). \tag{3.15}$$

where $\delta_{n_z}^{m}$ is the relative phase between $+m$ and $-m$ terms, or simply the clocking of each term, and can be calculated through the relation $\delta_{n_z}^{m} = \tan^{-1}(\frac{\beta_{n_z}^{m}}{\alpha_{n_z}^{m}})$.

Figure 3.15 (a) shows two spokes patterns with dominant azimuthal frequencies of $m=5$ and $m=9$. Figure 3.15 (c) shows the superposition of these spokes patterns with different relative phases. As can be seen in Figure 3.15 (b), the fit coefficient plots of all three height maps are the same, but the relative phase between $m=5$ and $m=9$ gives rise to the different height maps in (c). In Figure 3.15 (d), we have calculated the phase between the $m=5$ and $m=9$ terms of the surface maps shown in (c) using $\delta_{n_z}^{m} = \tan^{-1}(\frac{\beta_{n_z}^{m}}{\alpha_{n_z}^{m}})$. From this example, it is clear, the fit coefficient maps are not sufficient to retrieve the correct height maps and that there is degeneracy in the coefficient values.

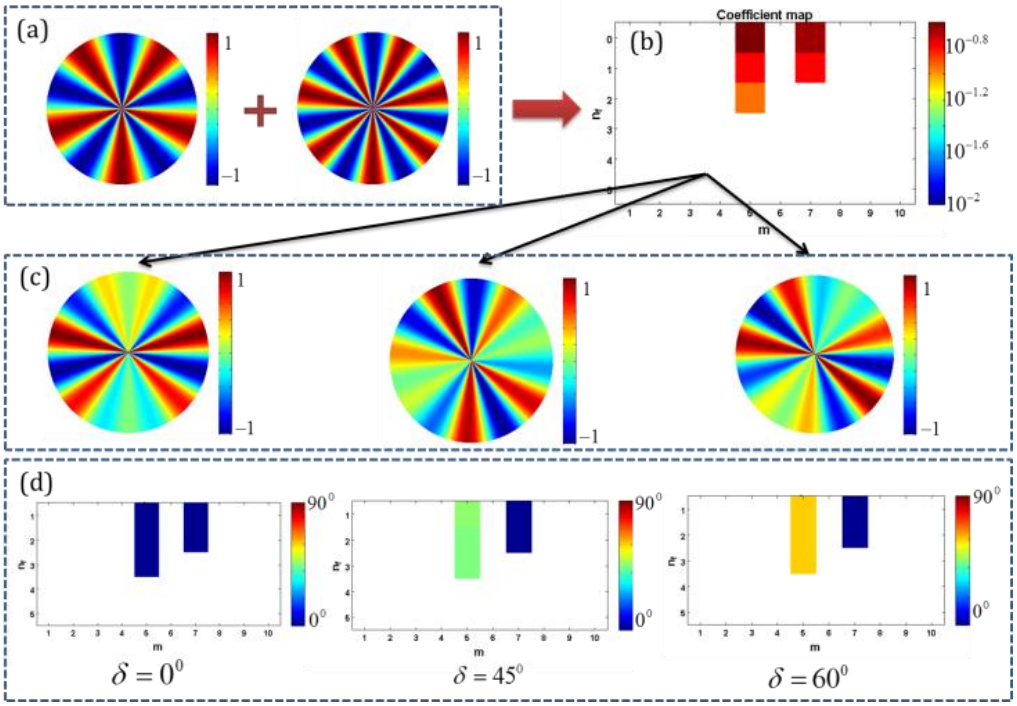


Figure 3.15 (a) Two spokes patterns with azimuthal frequencies of $m = 5$ and $m = 9$. (b) The resulting coefficient map of the combined map (superposition of maps in part (a)). (c) Different surface maps constructed by superposition of maps in part (a) with different relative phase. (d) Illustration of the phase difference in surface maps of part (c).

Applying the same ideas to the surface height map of Figure 3.9 (a) provides the phase representation, as shown in Figure 3.16. Due to the symmetry properties of Zernike polynomials, rotation of a $Z_{n_z}^{\pm m}$ by an angle $\frac{2\pi}{m}$ (clockwise or counter-clockwise), provides the same polynomial map. Therefore, in order to find the smallest angle between azimuthal orders of m, we have also calculated $\delta_{eff} = rem(\delta_{n_z}^m, \frac{2\pi}{m})$ where rem is the remainder after division (shown in Figure 3.16 (b)). The larger values of m have higher azimuthal symmetry and therefore δ_{eff} becomes smaller. The larger angles show more of positive or $\cos(m\varphi)$ terms and smaller angles show more of negative or $\sin(m\varphi)$ terms. To have meaningful results, for coefficients that are smaller than one-hundredth of a nanometer, the phase is set to zero to avoid small values in the denominator. This value is

below the noise level in the interferometer. For complicated textures like the one in Figure 3.9 (a), the coefficient phase map appears to be of limited use. However, looking at the selected azimuthal bands of the same surface as shown in Figure 3.10, it is obvious that the phase of neighboring azimuthal orders has an effect on the height map and overall RMS of the surface.

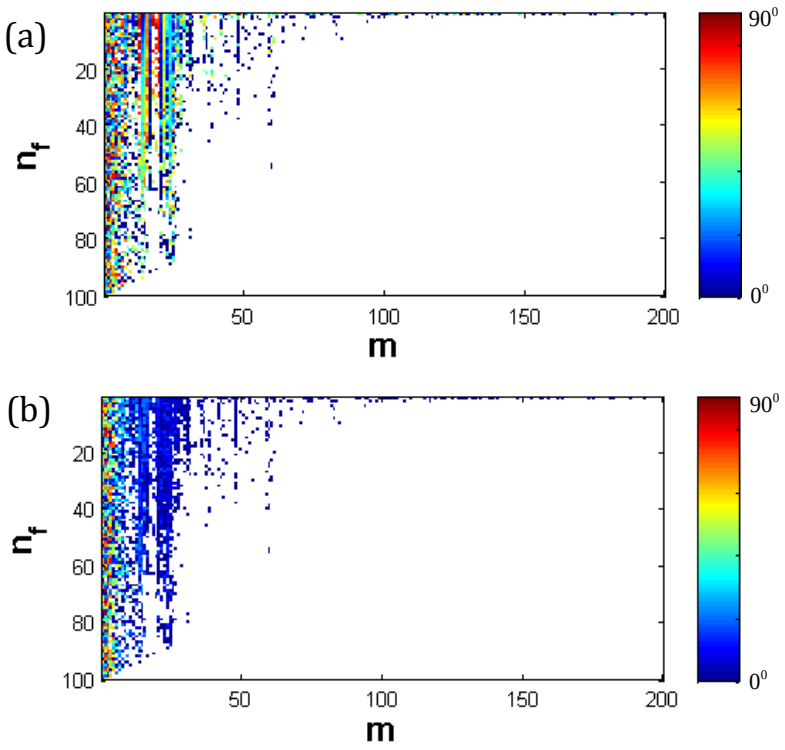


Figure 3.16 Illustration of phase for fit coefficients of the surface map in Figure 3.9 (a) For absolute phase calculated as $\delta_{n_z}^m = \tan^{-1}(\frac{\beta_{n_z}^m}{\alpha_{n_z}^m})$, and (b) relative phase calculated as

$$\delta_{eff} = rem(\delta_{n_z}^m, \frac{2\pi}{m}).$$

3.5 Convergence of fit residuals

The traditional understanding is that the least square fitting methods are ill-conditioned especially for calculating large order coefficients of Zernike polynomials. J. Y. Wang and D. E. Silva have shown that least square matrix inversion methods provide numerically

stable results despite the conventional belief [109]. They have shown that both methods, the Gram-Schmidt method, and the least-squares method give similar results. If the surface is not sampled properly, they will both become unstable [109]. It is common to sample the dataset and increase the number of samples as the number of basis elements goes up. For a reasonable fit, the number of samples N and the highest order of polynomials in the fit n are related to each other through $N \approx 9 \times n^2$ [93]. For example in [93], Kaya *et al.* increased the number of samples from 226 to 54845 as the number of Zernike polynomial terms goes up from 19 to 3319. They compared Zernike polynomials and Forbes Q-polynomial in terms of describing a freeform surface shape. They showed that using an edge-clustered grid will improve the efficiency of the fit for both polynomials and that Zernike and Q-polynomials are equally effective in describing freefrom shapes when placed on a base conic section [93].

We have kept track of the RMS of fit residuals as fitting to a higher number of terms. As the number of basis elements goes up, the RMS of fit residuals becomes smaller, and the rate at which the RMS of fit residuals drops reflects how well the Zernike representation isolates or captures the MSF texture or how much of the surface is comprised of form and MSF errors. In general, we see that the residual drops with increasing fit order rather slowly. As the number of terms is increased to capture more and more completely the MSF texture, eventually the Nyquist sampling limit near the edge of the aperture is violated. That is, the zero crossings of the Zernike polynomial at the aperture edge become more closely spaced than the pixels in the data. For a Zernike polynomial of order n_z and m, there are $(n_z - m)/2$ distinct roots in $0 < \rho < 1$ region [110]. For example, for Z_{200}^4, there should be 98 roots in $0 < \rho < 1$ whereas for the generated polynomial using a 1000x1000 uniform

grid, there are 93 zero crossings and five roots are already missing as shown in Figure 3.17. This impacts orthogonality and is discussed below. Interpolating our data to a non-uniform grid such as an edge-clustered [107] would enable robust fitting to even higher orders of polynomials. The RMS fit residual for 5151 basis elements (terms) for the germanium sphere in Figure 3.13 (a) is 13.6 nm. The RMS fit residual of the same surface for 20301 basis elements is 10.4 nm. This shows a 1% drop in RMS fit residual for an increase of 400 % in the number of basis elements. The RMS drops further to 7.8 nm after removing 45451 terms. The results of the RMS fit residual and the fit residual maps for the germanium sphere data are shown in Figure 3.18.

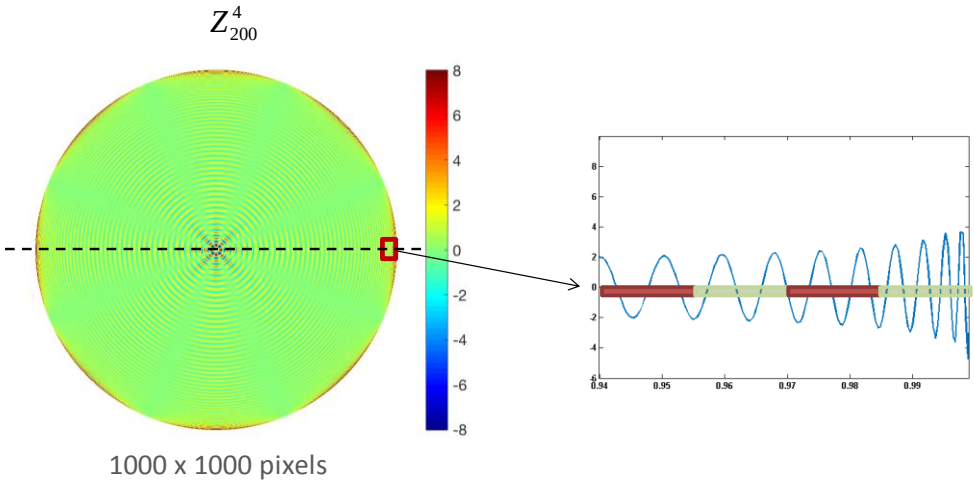


Figure 3.17 A large order Zernike polynomial Z_{200}^{4}. In this polynomial, the number of zero crossings between $0<\rho<1$ for a continuous case is 98. For a limited sized data such as 1000x1000 pixels, there are only 93 zero crossings in $0<\rho<1$.

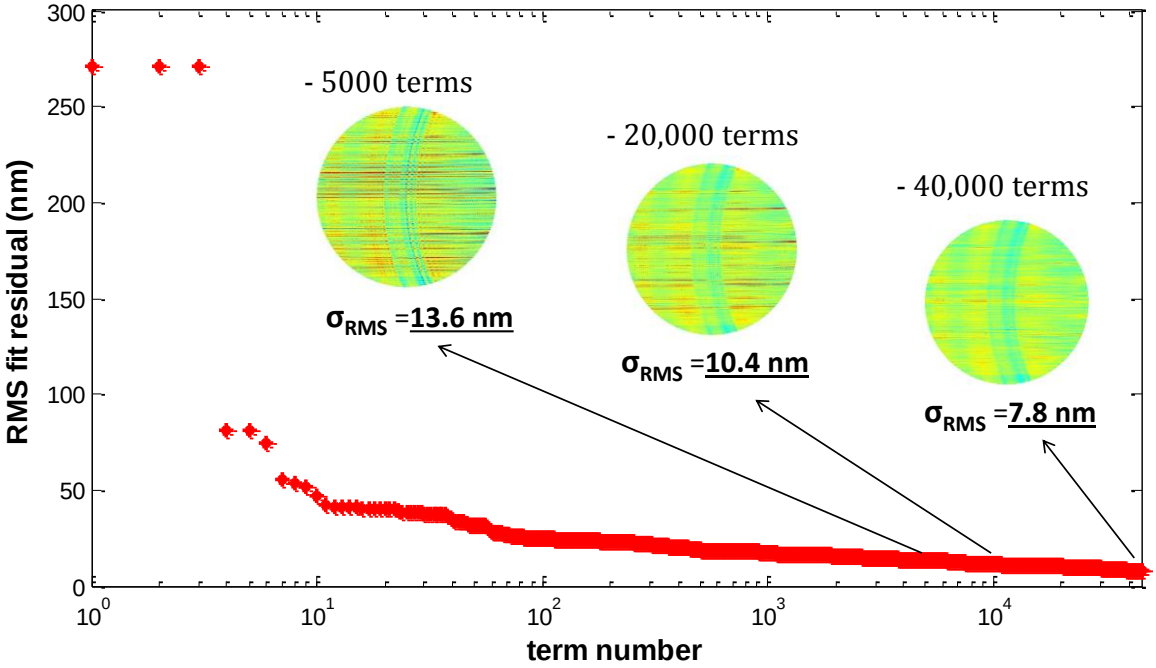


Figure 3.18 The evolution of RMS fit residual *vs.* number of basis elements. The RMS fit residual for 5151 basis elements is 13.6 nm, for 20301 basis elements is 10.4 nm, and for 45451 terms is 7.8 nm.

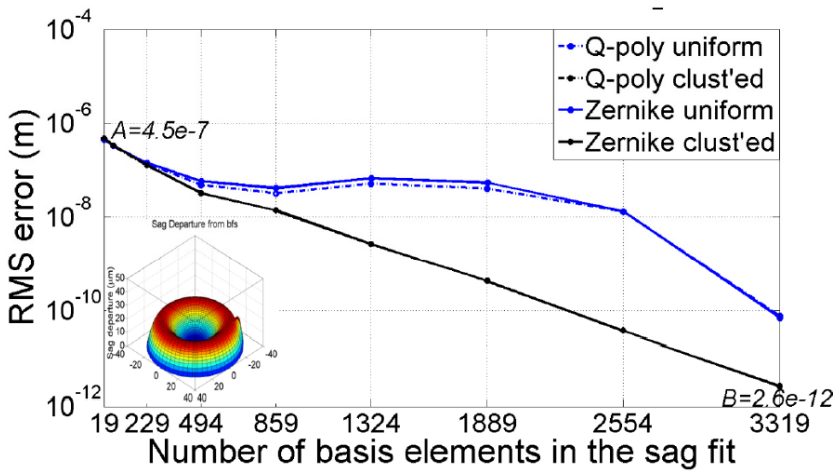


Figure 3.19 Uniform and Edge clustered sampling in fitting a parabola with a bump at the edge to Zernike polynomials and Forbes Q-polynomials [93].

Zernike polynomials are orthogonal to arbitrary large orders if the datasets are continuous. Since we are dealing with discrete datasets and the surface height maps are

limited to a certain number of pixels, an approximation to orthogonality breaks down at extremely large orders. One consequence of orthogonality is that the fit coefficients remain the same regardless of the number of terms being used for the fit. One way to check this condition is to calculate the RSS of the fit coefficients plus the square of RMS residual map (σ_{res}) i.e. $\sigma_{fit} = \sqrt{\sum_n \sum_m (a_{n_z}^m)^2 + (\sigma_{res})^2}$ as a function of the number of terms in the fit. This value should remain constant and should be equal to the RMS of the surface as shown in Figure 3.20 for an orthogonal representation. This quantity is reasonably constant in our analysis with deviations of less than 1 nm, suggesting orthogonality is reasonably well maintained over much of our fitting range. In Figure 3.20, the difference between σ_{fit} and the RMS of the surface stabilizes after about 90 terms fit and then increases after about 20,000 terms. The fact that this difference is not a constant is a consequence of non-orthogonality over the discrete dataset. The increase of the difference between σ_{fit} and the RMS of the surface after stabilizing may be the onset of a significant breakdown of orthogonality as a consequence of low data density compared to the spacing between adjacent zero crossings.

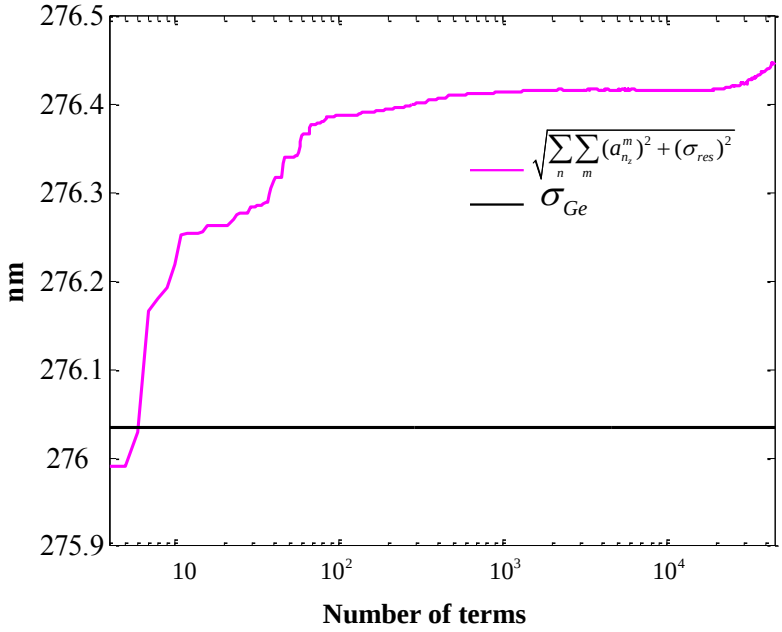


Figure 3.20 Check for orthogonality of polynomials up to 45451 terms. The difference between σ_{fit} (pink curve) and the RMS of the surface is less than a nanometer across 45451 terms.

The Zernike polynomials generated using the recursion formulas in this work extend to more than 45,000 terms and represent a unique possibility for characterization of MSF on freeform surfaces. We discussed how well Zernike polynomials describe different aspects of MSF structures such as type (raster, ring-like, etc.), orientation, length scale of the period, magnitude of different length scales and their relative phase. Although extracting information about length scale and amplitude is possible through PSD, it involves slicing through the two-dimensional PSD or integrating the two-dimensional PSD into a one-dimensional PSD in addition to the pre-processing steps such as windowing and detrending, hence losing information on the low order form errors and anisotropy of MSF texture. Unlike area PSD where we can see all the frequency peaks in the data, the polynomial representation is limited to the number of orders and will limit capturing the high-frequency content of the surface depending on the number of basis elements used. For

example, in the germanium sphere using area PSD, there is a peak at $27(1/mm)$. This corresponds to 216 cycles across the aperture and in order to capture that frequency with Zernike polynomials, 340 orders would be required. However, unlike Fourier methods, polynomial coefficient distributions provide a different insight into the signatures of MSF texture. Such information can be useful in the specification and predicting the effects of MSF on optical performance. We have observed that Zernike and Forbes Q-polynomials have similar characteristics in representing mid-spatial frequency content of the surface. The RSS of Zernike-fit coefficients give access to RMS wavefront error that is directly related to the Strehl ratio [111] and RSS of Q-polynomial coefficients provide the RMS gradient of the wavefront that is related to the ray transverse error [28]. The key to performing a robust fitting to both polynomials is the use of recurrence relations [87]. Polynomial fitting methods open the door to spatial frequency filtering and isolation of specific MSF textures with different orientations caused by different fabrication processes, and they provide an interesting tool for quantifying and specifying MSF in manufacturing, design, and conformance testing.

CHAPTER 4 : OPTICAL PERFORMANCE

In this chapter, we introduce optical performance metrics such as Strehl ratio and modulation transfer function (MTF) and will show examples of those metrics when there is MSF texture on the surface. The goal is to have an ideal analysis method to isolate the surface data so that designers can incorporate it in their tolerancing. This will then allow for a clear quantitative specification, where the designer has confidence that meeting a specification threshold will guarantee adequate optical performance, and the fabricator making the surface has a way of measuring and applying a conformance test to the part. We will investigate how Zernike polynomials can be used for tolerancing of mid-spatial frequencies in the next chapter.

All optics should be toleranced before manufacturing. The tolerancing should guarantee the performance requirement and minimize the cost of manufacturing at the same time, and these two are usually contradictory [14]. In general, performance metrics are application specific and can be grouped based on the fields of illumination, imaging or beam shaping. Common metrics for optical performance in imaging systems are RMS spot size, encircled energy and Strehl ratio. MTF can be used by specifying a certain MTF value at a specific spatial frequency or over a specific frequency band. J. Tamkin recommends using the maximum drop in MTF from the nominal performance for specifying MSF surfaces [15, 24]. Different frequency components of the surface have different scattering characteristics. Form errors do not affect the nodes of point spread function, they only affect Strehl ratio

and reduce the peak intensity. High spatial frequency or roughness, does not change the distribution of light between different orders, but rather reduces the peak of all orders equally. Mid-spatial frequency errors are grating-like structures and affect the distribution of light into different orders. These effects can be dramatic and significantly impact performance. Examples are shown in Figures 4.1 to 4.3.

The optical reflection or transmission characteristics of a surface are directly impacted by the surface geometry, and they are a direct measure of optical performance. In the limit of a relatively smooth surface that is stationary, a clear relationship exists between measured scatter and the power spectral density (PSD) of the surface and hence surface topographical features [36]. These assumptions are not valid for surfaces with MSF texture [36].

In imaging applications, the modulation transfer function (MTF) is the most common way to predict image quality. The MTF provides insight into the spatial response of an imaging system. For surfaces with significant MSF texture, the RMS or Strehl ratio metric may not capture the performance degradation, and the more detailed information contained in an MTF is likely the more suitable approach [15, 24]. Marioge *et al.* have applied a rotational periodic MSF structure to a surface near the pupil. The resulting MTF plot in the image plane shows oscillations. They have used the autocorrelation function to obtain these curves [112]. S. Shikama is one of the first authors to investigate the effects of structured spatial frequencies on the image plane. He has used the commercially available software ZEMAX for his simulations, which is based on geometrical ray tracing [113]. In the following sections, we will analyze the impact of MSF texture on the point spread function, the Strehl ratio and the MTF response of the system. Our analysis extends this work and

performs these calculations on a range of measured and virtual surfaces, using MATLAB and the commercially available software MetroProX (MX v6.3.0.4).

4.1 Point spread function, modulation transfer function, and Strehl ratio

The point spread function (PSF), or the impulse response, indicates the characteristics of the image of a point source [25]. The wider the PSF, the poorer the imaging resolution. Asymmetry in PSF causes image distortion. PSF of a system is impacted by both diffraction and aberration.

The modulation transfer function is the magnitude response of the optical system to different spatial frequencies [114]. In other words, it is the magnitude of the optical transfer function (OTF) of the system and is a measure of how well the system can resolve different spatial frequencies. The coherent transfer function (amplitude transfer function in [25]) is the Fourier transform of the point spread function (amplitude impulse response),

$$H(f_u, f_v) = \int_{-\infty}^{\infty}\int h(u,v)\exp[-j2\pi(f_x u + f_y v)]dudv\,. \tag{4.1}$$

The incoherent transfer function is the normalized autocorrelation function of the amplitude transfer function defined as [25]

$$\mathbb{H}(f_u, f_v) = \frac{\int_{-\infty}^{\infty}\int |h(u,v)|^2 \exp[-j2\pi(f_x u + f_y v)]dudv}{\int_{-\infty}^{\infty}\int |h(u,v)|^2\, dudv}. \tag{4.2}$$

Therefore,

$$\mathbb{H}(f_u, f_v) = \frac{\Im\{|h(u,v)|^2\}}{\int_{-\infty}^{\infty}\int |h(u,v)|^2\, dudv}, \tag{4.3}$$

where $h(u,v)$ is the PSF.

Most MTF plots contain the diffraction-limited MTF line for reference and MTF graphs for several field points. The OTF of a diffraction limited system with a circular pupil $P(x,y) = circ(\frac{\sqrt{x^2+y^2}}{w_{xp}})$ for incoherent illumination is

$$\mathbb{H}(\rho) = \begin{cases} \frac{2}{\pi}\left[\arccos(\frac{\rho}{2\rho_0}) - \frac{\rho}{2\rho_0}\sqrt{1-(\frac{\rho}{2\rho_0})^2}\right] & \rho \le 2\rho_0 \\ 0 & otherwise \end{cases}, \tag{4.4}$$

where

$$\rho_0 = \frac{w_{xp}}{\lambda z_{xp}}. \tag{4.5}$$

MTF is easy to measure and the aggregate MTF of several subsystems is the multiplication of their individual MTFs if each subsystem operates independently [114].

Strehl ratio is another common performance metric for imaging systems and is the ratio of the central irradiance (or PSF) for an aberrated pupil to its value for an unaberrated pupil (Eqn. 4.6). The Strehl ratio is also equal to the ratio of the area under the aberrated MTF to the corresponding area for an aberration-free system (Eqn. 4.7). Over circular apertures, the Strehl ratio can be defined by Eqn. 4.8. A good approximation for the Strehl ratio is given by Eqn. (4.9), and is valid for a Strehl ratio as small as 0.1 [29].

$$S = \frac{I_{ab}}{I_{unab}} = \left.\frac{PSF_{ab}}{PSF_{unab}}\right|_{(0,0)} \tag{4.6}$$

$$S = \frac{\iint_{f_u, f_v} \left[MTF_{ab}\right] dudv}{\iint_{f_u, f_v} \left[MTF_{unab}\right] dudv} \tag{4.7}$$

$$S = \frac{1}{2\pi}\left|\int_0^1\int_0^{2\pi} e^{ik\Delta W(r,\theta)} r dr d\theta\right|^2 \tag{4.8}$$

$$S \approx \exp(-k^2\sigma^2) \tag{4.9}$$

Figures 4.1 to 4.3 show the PSF and MTF of different spatial frequency components of the IRG26 sphere shown in Figure 1.3. We have decomposed the surface into different spatial frequency components: spatial periods of $\Lambda > 3mm$ (form errors), $0.1mm < \Lambda < 3mm$ (mid-spatial frequency region), and $\Lambda < 0.1mm$ (high-spatial frequency region). For each spatial frequency range, the 2D PSF and the 2D MTF of the corresponding surface were generated using the commercial software MetroProX (Mx). Since anisotropy is present in most cases of MSF texture, a 2D MTF analysis where one can see the optical performance in all directions is important. The diffraction analysis in Mx enables the user to qualify the optical surface at a single field point for imaging performance through a circular exit pupil, assuming a transmissive wavefront. This analysis is performed using the FFT operation upon a measured wavefront. If not selected otherwise, the software assumes an exit pupil with $f/\# = 1$, NA of 0.5, a wavelength of 0.632 μm, and a refractive index of 1.51. In Figure 4.1 (c), one can see the diffracted orders in the PSF due to the grating-like characteristics of the MSF error on the pupil function. The Strehl ratio for this MSF content of the wavefront is 0.85. The Strehl ratio for the isolated roughness region is 0.99 and for form errors of the same wavefront is 0.92.

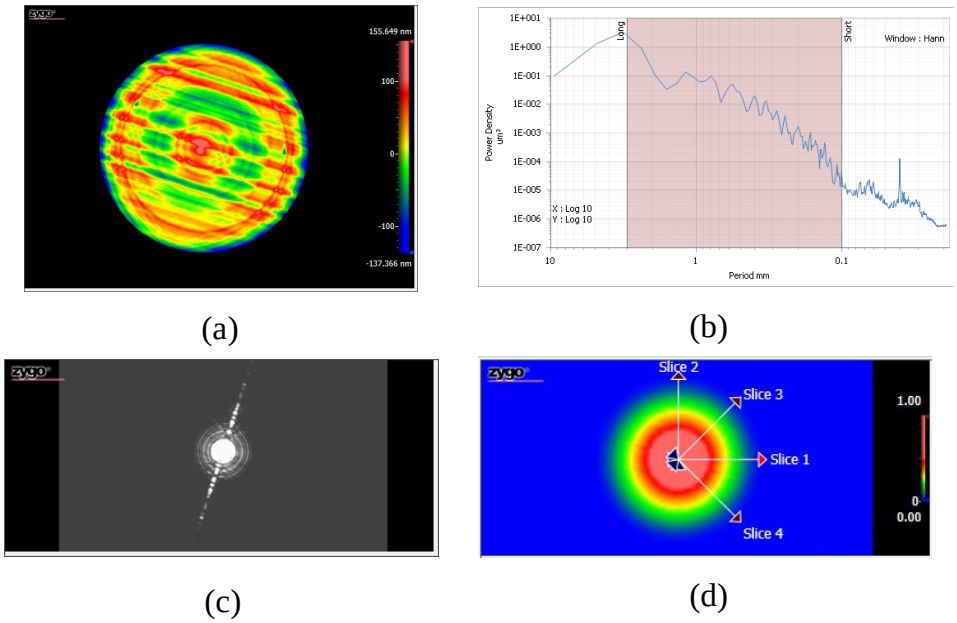


Figure 4.1 (a) Mid-spatial frequency content of IRG 26 sphere (spatial periods of $0.1mm < \Lambda < 3mm$). (b) PSD of the whole surface with the selected spatial bandwidth marked in the red box. (c) 2D PSF of the surface shown in (a). (d) 2D MTF of the surface shown in (a). The Strehl ratio is 0.85.

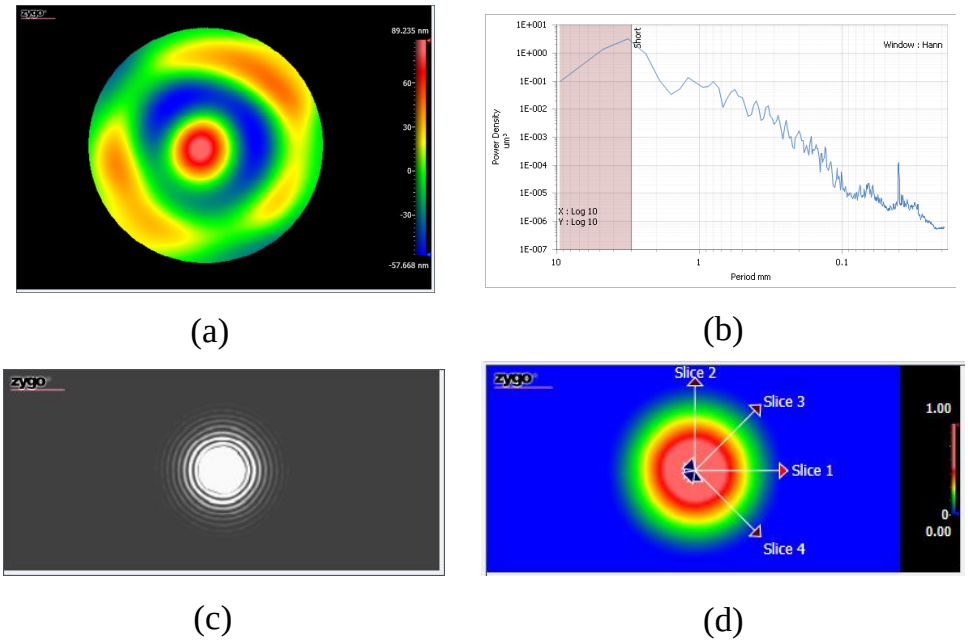


Figure 4.2 (a) Low spatial frequency content of IRG 26 sphere (spatial periods of $\Lambda > 3mm$). (b) PSD of the whole surface with the selected spatial bandwidth marked in the red box. (c) 2D PSF of the surface shown in (a). (d) 2D MTF of the form errors. The Strehl ratio is 0.92.

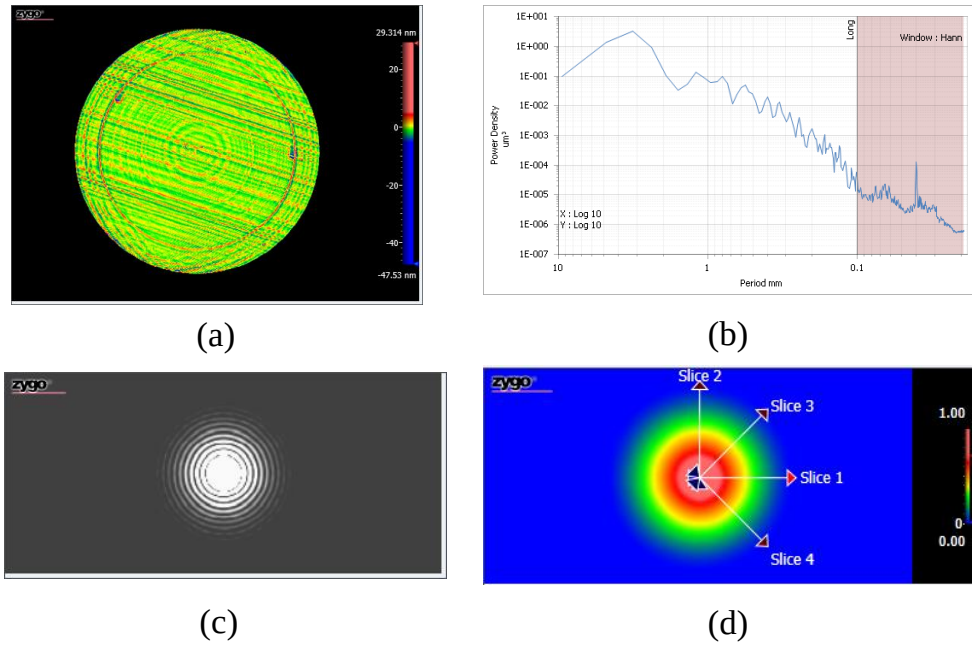


Figure 4.3 (a) High-frequency content of IRG 26 sphere (spatial periods of $\Lambda < 0.1mm$). (b) PSD of the whole surface with the selected spatial bandwidth marked in the red box. (c) 2D PSF of the surface shown in (a). (d) 2D MTF of the surface shown in (a). The Strehl ratio is 0.99 that indicates the small effect of roughness on optical performance.

4.2 Degradation due to specific MSF signatures

In this section, you can see the MTF degradation due to typical MSF structure on optical surfaces. Figure 4.4 shows a sketch of an aberrated wavefront error at the exit pupil. The wavefront error can be described as the difference between the aberrated wavefront and the spherical (ideal) wavefront as in Eqn. 4.10,

$$W(x,y) = W_{ab}(x,y) - W_{ideal}(x,y). \tag{4.10}$$

The aberrated pupil function is defined by Eqn. 4.11 where the circ function is the circular diffraction limited exit pupil and w_{xp} is the exit pupil radius.

$$P(x,y) = circ\left(\frac{\sqrt{x^2+y^2}}{w_{xp}}\right)\exp\left(-jkW\left(x/w_{xp}, y/w_{xp}\right)\right) \tag{4.11}$$

The simulated optical systems using MATLAB in this dissertation have a diffraction-

limited performance and operate at $f/\#=5$. We also assume that the beam footprint covers the lens and the exit pupil and entrance pupil are the same and are equal to the diameter of the lens. The wavefront error with mid-spatial frequencies is then added to the ideal spherical wavefront. The cut-off frequency for this system is 316 cycles/mm, assuming the incident wavelength is in the visible range ($\lambda = 633$ nm).

The cut-off frequency is the limit where spatial frequencies greater than the cut-off cannot be resolved by the lens. The cut-off frequency depends on the size of the exit pupil, the focal length of the lens and the incident wavelength (see Eqn. 4.5).

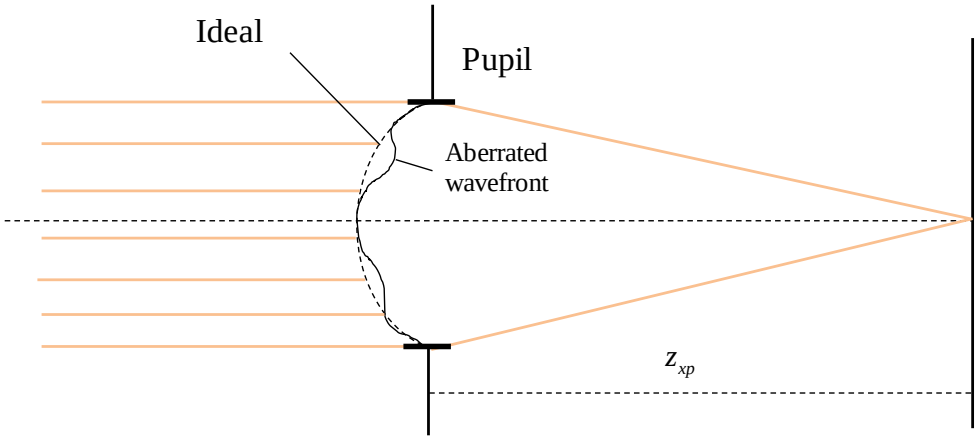


Figure 4.4 Sketch of the exit pupil with aberrated and ideal wavefronts.

As mentioned earlier, MSF patterns are grating-like structures. Investigating the optical performance of synthetic periodic structures will provide insight into how to interpret the optical performance impact of real MSF texture. Figure 4.5 (a) shows raster-like patterns of a sinusoidal surface with the PV of 0.3 waves and 12 cycles across the aperture. We have generated the resulting 2D MTF shown in Figure 4.5 (b) by inserting the wavefront in (a) on the exit pupil and propagating the wavefront using FFT methods. A horizontal and vertical cross sections of the 2D MTF plot in part (b) are shown in Figure 4.5 (c) in comparison with diffraction limited MTF. As can be seen, the horizontal cross section of

2D MTF has an oscillating behavior whereas the vertical cross section of 2D MTF is the same as diffraction limited MTF. The number of oscillations in the horizontal MTF cross section is equal to the number of cycles across the aperture. The horizontal and vertical MTF cross sections are labeled as uMTF and vMTF respectively.

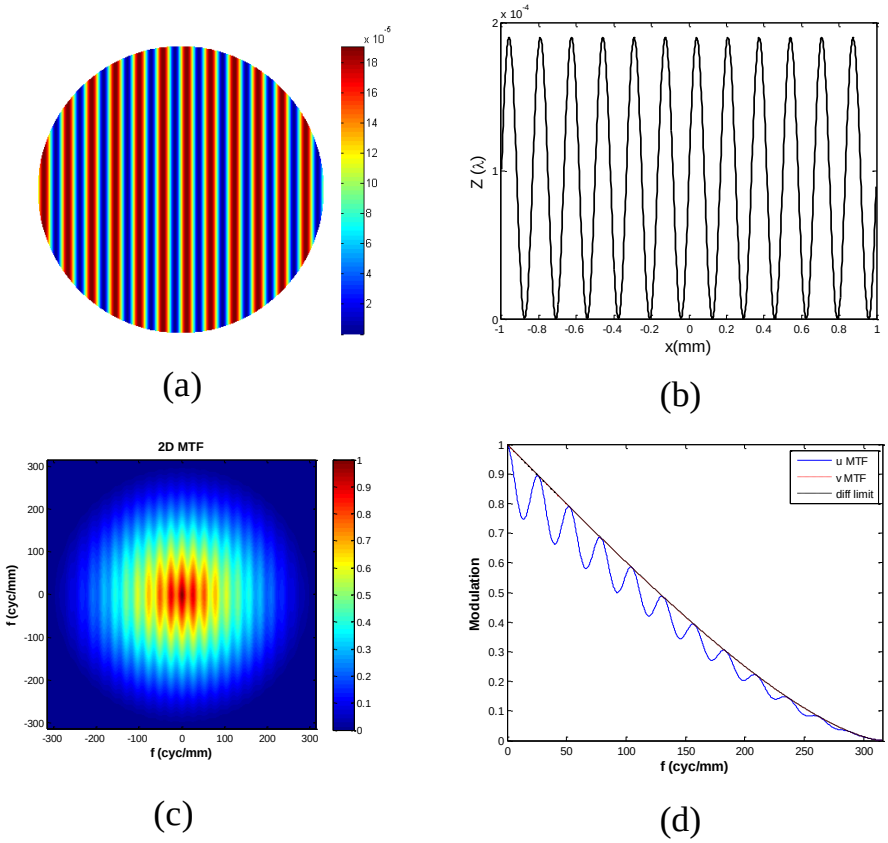


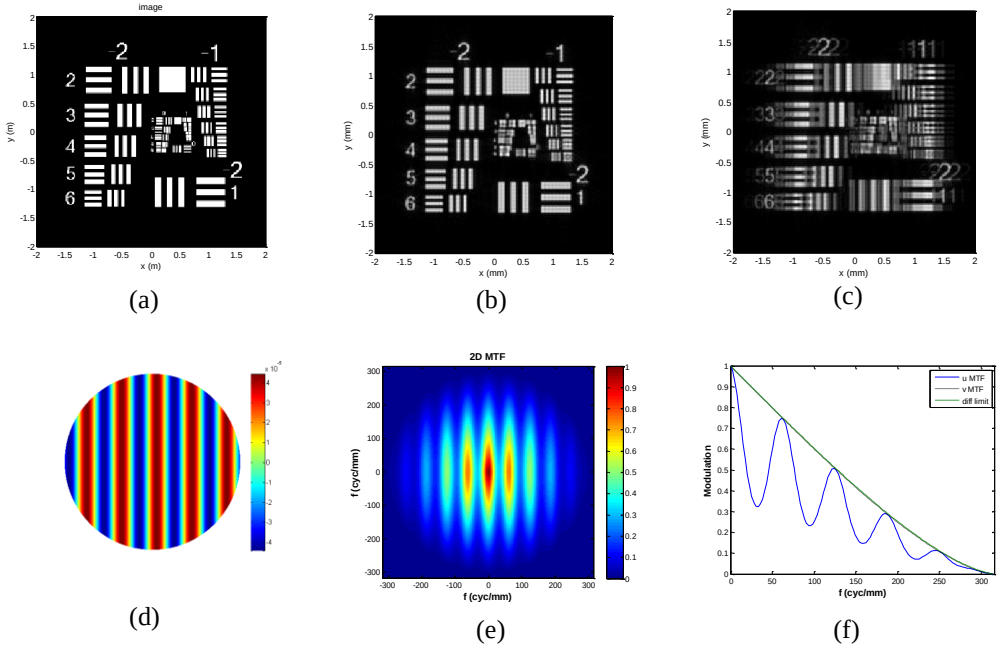
Figure 4.5 (a) Sinusoidal raster patterns with a PV of 0.3λ (b) Cross section of the sinusoidal raster-like patterns. (C) 2D MTF of the system with sinusoidal MSF texture on the exit pupil. (c) Vertical and horizontal cross sections of 2D MTF. The normalized area MTF drop from the diffraction-limited MTF is 11%. The peak MTF drop and average MTF drop from the diffraction-limited MTF are 0.2 and 0.021, respectively.

To see the effect of MSF texture on an image, we have simulated the image of USAF 1951 resolution chart with and without MSF texture on the exit pupil. The USAF test chart was obtained from reference [115]. Figure 4.6 (a) and Figure 4.7 (a) show the ideal image of USFA chart, Figure 4.6 (b) and Figure 4.7 (b) show the image when there are no

aberrations on the exit pupil and part (c) in both figures shows the image when we have sinusoidal raster-like MSF texture on the exit pupil. Vertical raster patterns will result in ghost images in the horizontal direction -Figure 4.6 (c)- and horizontal raster patterns will result in vertical ghost images -Figure 4.7 (c). Part (e) in both figures shows the 2D MTF and part (f) shows the cross-sectional MTFs. The approach for simulating these images are based on Eqn. 4.12. I.e. to simulate the final image we take the inverse Fourier transform of the product of the transfer function and the ideal image.

$$U_i(u,v) = \Im^{-1}\left\{H(f_U, f_V)\Im\left\{U_g(u,v)\right\}\right\}. \tag{4.12}$$

In Eqn. 4.12, $H(f_U, f_V)$ is the coherent transfer function and $U_g(u,v)$ is the ideal image.



(a) (b) (c)

(d) (e) (f)

Figure 4.6 (a) USAF test chart (ideal image) (b) Diffraction-limited image with no aberration on the wavefront (c) Aberrated image with mid-spatial frequency texture on the exit pupil. (d) Vertical raster-like MSF texture on the exit pupil (e) 2D MTF of the imaging system with vertical raster patterns on the exit pupil. (f) Horizontal and vertical cross sections of the 2D MTF.

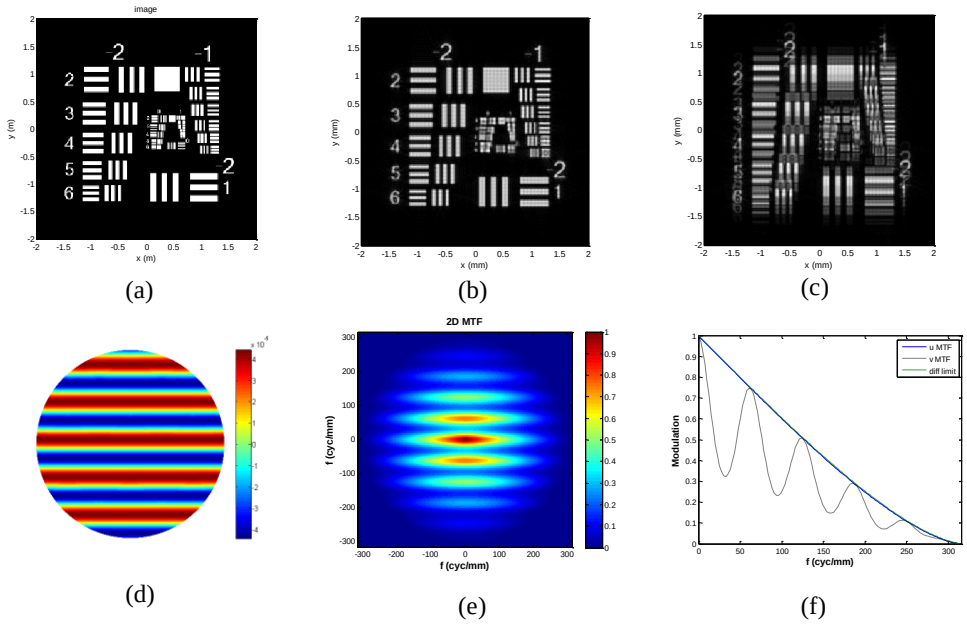


Figure 4.7 (a) USAF test chart (ideal image) (b) Diffraction-limited image with no aberration on the wavefront (c) Aberrated image with mid-spatial frequency texture on the exit pupil (d) Horizontal raster-like MSF texture on the exit pupil (e) 2D MTF of the imaging system with horizontal raster patterns on the exit pupil. (f) Horizontal and vertical cross sections of the 2D MTF.

In addition to sinusoidal patterns, it is also worthwhile to examine the optical performance and PSD of cusp-shaped surfaces, since the residual tool marks of diamond turned surfaces are cusp-shaped and not purely sinusoidal [36]. H. Aryan *et al.* have performed an in-depth study to connect design specifications, acceptance test criteria and optical performance of freeform optics with residual surface errors [116].

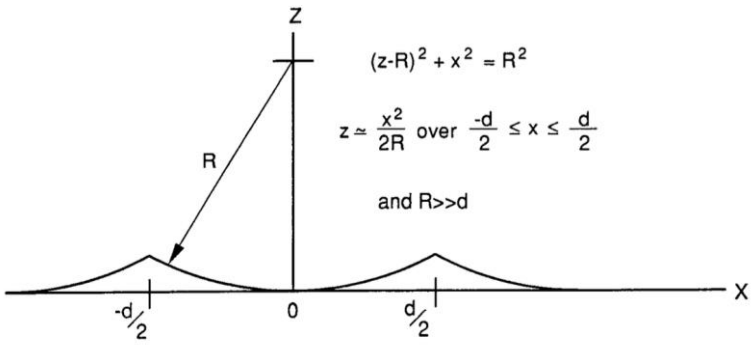


Figure 4.8 Sketch of a cusp-shaped surface generated by a tool of radius R and feed rate *d* [36].

Figures 4.9 shows a cusp-shaped surface with the same periodicity and PV as Figure 4.5. Although they have the same period and PV, their optical performance is significantly different. The drop in the area MTF curve from its un-aberrated limit for the sinusoidal structure in Figure 4.5 is 38% whereas for the cusp-shaped surface in Figure 4.9 the area MTF drop is only 7.8%. The peak MTF drop for the wavefront in Figure 4.5 is 0.67 and the average MTF drop from its diffraction-limited curve is 0.073. The peak MTF drop for the cusp-shaped surface in Figure 4.9 is 0.14 and the average MTF drop from its diffraction-limited curve is 0.015. Like Figure 4.5, the horizontal and vertical cross sections of the 2D MTF have different behavior due to the anisotropy of the MSF texture. The horizontal MTF (uMTF)- blue line in Figure 4.9 (c)- oscillates with the same number of wiggles as the MSF texture. The vertical MTF (vMTF) is the same as diffraction limited MTF.

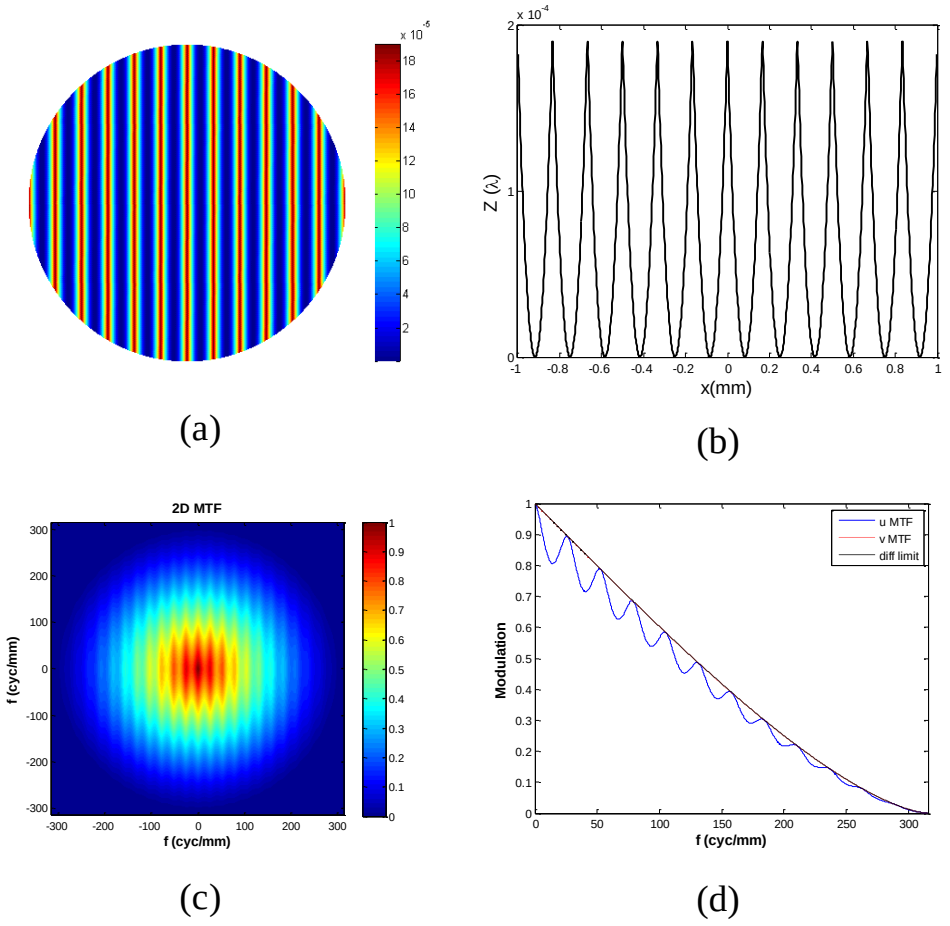


Figure 4.9 (a) Cusp shaped raster pattern with a PV of 0.3λ (b) 2D MTF of the system with cusp-shaped MSF texture on the exit pupil. (c) Vertical and horizontal cross sections of the 2D MTF. The area MTF drop is 7.8% for this cusp-shaped surface. The peak MTF drop and average MTF from the diffraction limited MTF curve are 0.14 and 0.015, respectively.

The PSD plots of the two surfaces in Figures 4.5 and 4.9 depict the difference in their frequency content despite the fact that they have the same period and same PV. A pure sinusoidal grating diffracts light into different orders and their locations in the image plane are described by the well-known grating equation. A cusp-shaped surface diffracts light into multiple spots but their locations in the image plane are not straight forward. This is due to the fact that a cusp-shaped surface can be generated from a superposition of multiple sinusoidal waves of different amplitudes. Looking at the PSD curve of a pure sinusoidal

wave, one can see an obvious peak whereas the PSD of the cusp-shaped surface shows multiple peaks with lower values as shown in Figure 4.10. The RMS height for sinusoidal surface is $\sigma = \frac{PV}{\pi} \approx 0.32$ PV and for the cusp-shaped surface is $\sigma \approx 0.25$ PV. We can see that PV and RMS of the surface are not ideal metrics for optical performance since they are not linearly proportional to the drop in MTF.

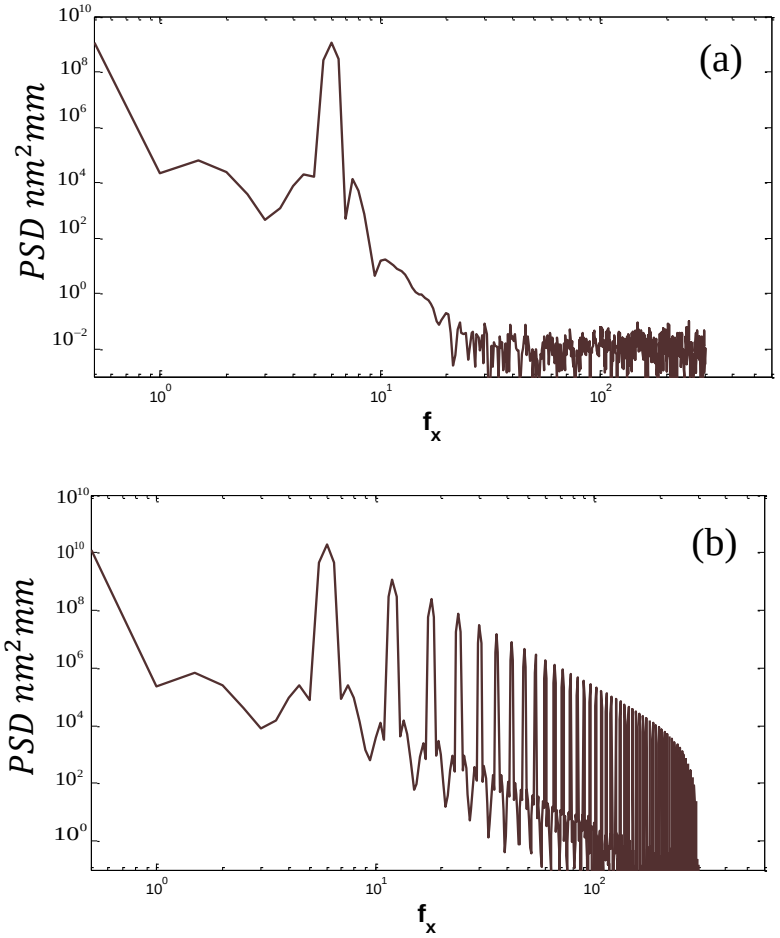


Figure 4.10 (a) PSD of sinusoidal raster pattern with 12 cycles across aperture and PV = 0.3λ. (b) PSD of the cusp-shaped surface with the same period and PV.

Figure 4.11 (a) shows a simulation of a radial sinusoidal pattern. Figure 4.11 (c) shows the 2D MTF and is obtained by inserting the radial sinusoidal wavefront on the exit pupil with a given $f_{\#}$. The resulting 2D MTF is rotationally-invariant and has the same

symmetry properties of the wavefront. The MTF behavior of radial periodic structures is different from the raster-like patterns. For radial patterns, there is a larger drop at lower frequencies as well as an overall drop in MTF. For a radial sinusoidal pattern with PV = 0.3λ and 12 cycles across the aperture, the area MTF drop is 38 % and the Strehl ratio is 0.62. The peak MTF drop is 0.47 and the average MTF drop from the diffraction limited MTF curve is 0.073. The maximum MTF drop is larger for radial patterns compared to raster patterns of the same PV, and a cross-sectional PSD will not show this difference. The average MTF drop from diffraction limited MTF curve is the same for radial and raster sinusoidal patterns of the same period and same PV.

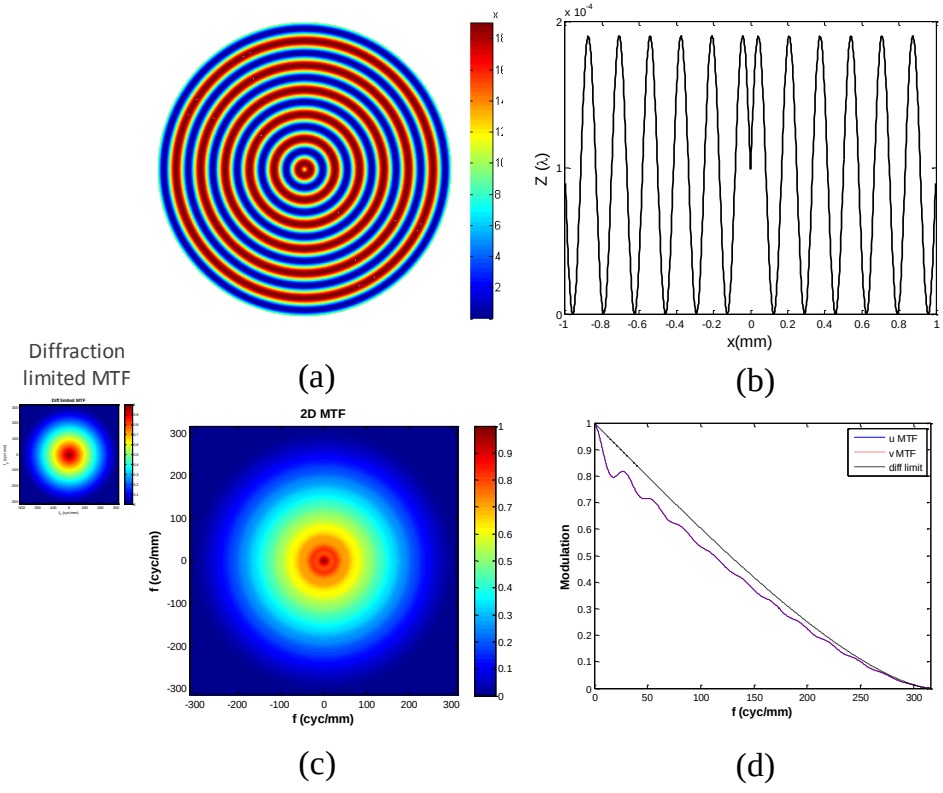


Figure 4.11 (a) Sinusoidal radial pattern with a PV of 0.3λ (b) Cross section of (a). (c) 2D MTF of the system with sinusoidal radial MSF texture on the exit pupil. The upper left image is the diffraction limited 2D MTF (d) Vertical and horizontal cross sections of the 2D MTF. The area MTF drop is about 11% and the Strehl ratio is 0.89. The peak MTF drop is 0.13 and the average MTF drop is 0.021, respectively.

Figure 4.12 shows the MTF response of a system with radial cusp-shaped texture on exit pupil. As we saw before, the MTF drop for the cusp-shaped surface is less than the sinusoidal texture of the same PV and the same period. The peak MTF drop and average MTF drop from its diffraction-limited curve are 0.098 and 0.015 respectively.

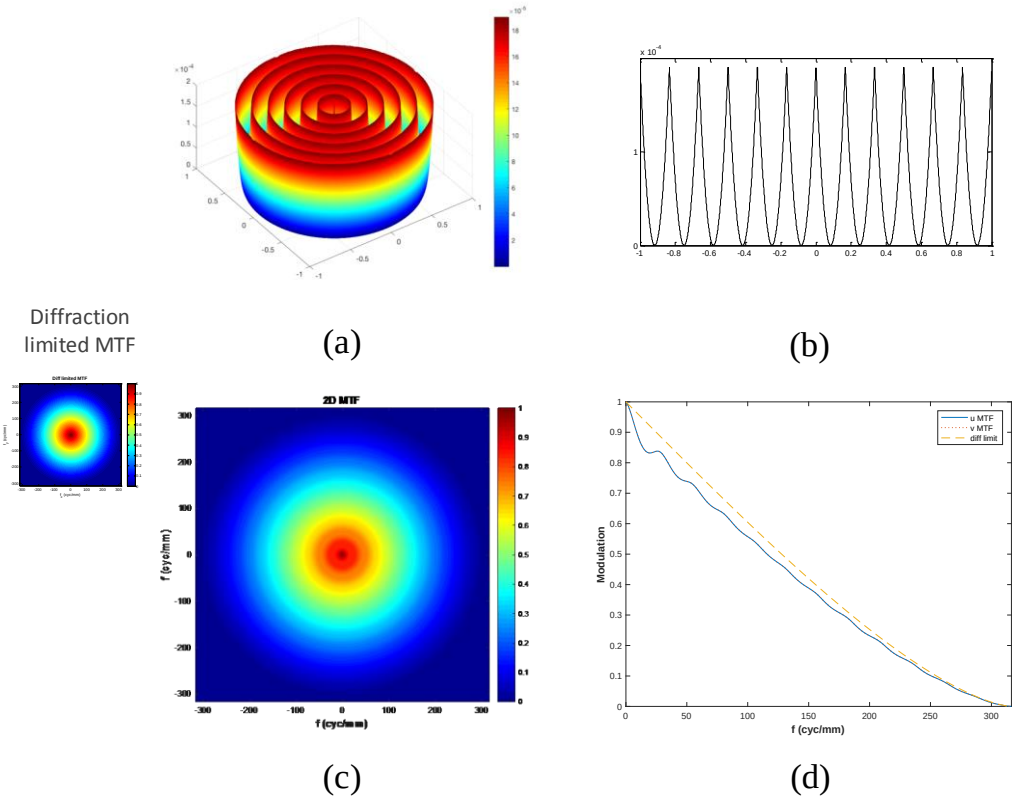


Figure 4.12 (a) A cusp-shaped radial pattern with a PV of 0.3λ (b) Cross section of the surface in part (a) (c) 2D MTF of the system with cusp-shaped radial MSF texture on the exit pupil. (d) Vertical and horizontal cross sections of 2D MTF. The area MTF drop is about 7.5% and Strehl ratio is 0.92. The peak MTF drop and average MTF drop are 0.098 and 0.015, respectively.

We have seen that the MTF behavior is different for different shapes of MSF texture. Since MSF textures are similar to diffraction gratings and the efficiency of a diffraction grating is a non-linear function of its PV, we also expect for MTF to change with the amplitude of MSF texture in a non-linear fashion. The relationship between MTF loss and PV of MSF texture has been shown by J. Tamkin in [15]. To explore this further, we have

carried out the MTF loss calculation for both sinusoidal and cusp-shaped radial patterns as a function of PV. The result is shown in Figure 4.13.

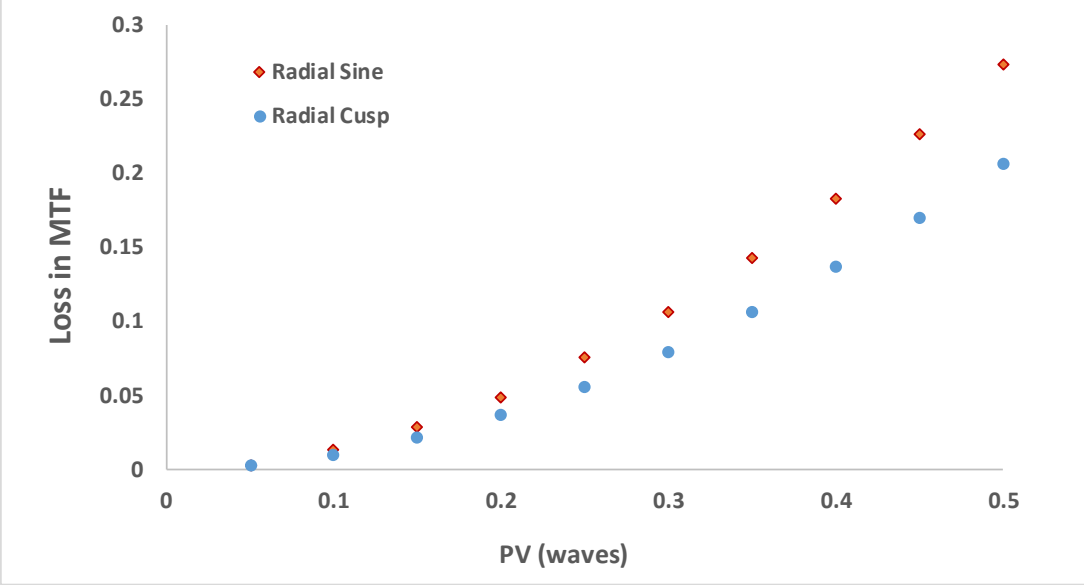


Figure 4.13 Area MTF loss *vs.* PV. Sinusoidal patterns have a larger loss compared to cusp-shaped patterns of the same PV.

Figure 4.13 confirms the previous observation that the MTF loss of a pure sinusoidal MSF texture is larger than the MTF loss a cusp-shaped surface with the same period and PV. In this simulation, surfaces are assumed to be transmissive with a refractive index of 1.51 and incident wavelength of 633 nm. The loss is defined as the Strehl ratio of the ideal system minus Strehl ratio of the system with MSF texture where Strahl ratio is obtained by using Eqns. 4.6 and 4.7.

The MTF loss depends on refractive index as well as PV. Figure 4.14 shows the loss in MTF for radial cusp surfaces with 12 cycles across aperture with the same PV but different materials with different refractive indices. In these simulations, we ignore the loss due to material absorption. For a larger refractive index, the MTF loss will be larger. This is expected since a larger refractive index with the same PV results in a larger OPD and a

larger wavefront error. Therefore, optical path difference (OPD) is a more suitable metric for optical performance compared to PV.

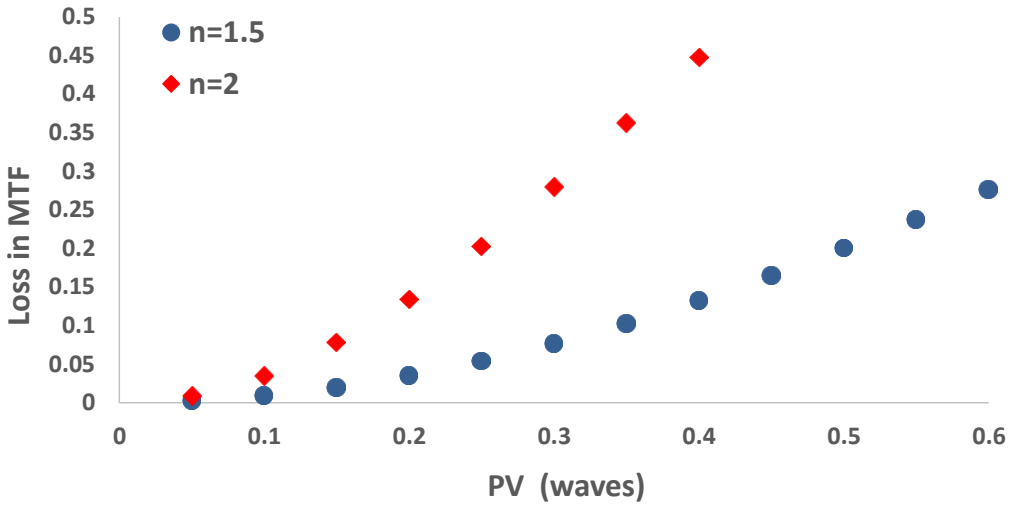


Figure 4.14 The MTF loss *vs.* PV has a higher slope for a larger refractive index.

If we plot the MTF loss as a function of OPD rather than PV, we can see the dependence of MTF loss to refractive index more clearly. In Figure 4.15, we have plotted the loss in MTF as a function of OPD for two different materials where OPD is equal to $(n-1)PV$. As can be seen, the loss in MTF *vs.* OPD is the same for different materials of the same PV. This statement is valid for transmission optics. For reflective optics, the optical path difference (OPD) does not depend on the refractive index but it is equal to twice the height variations.

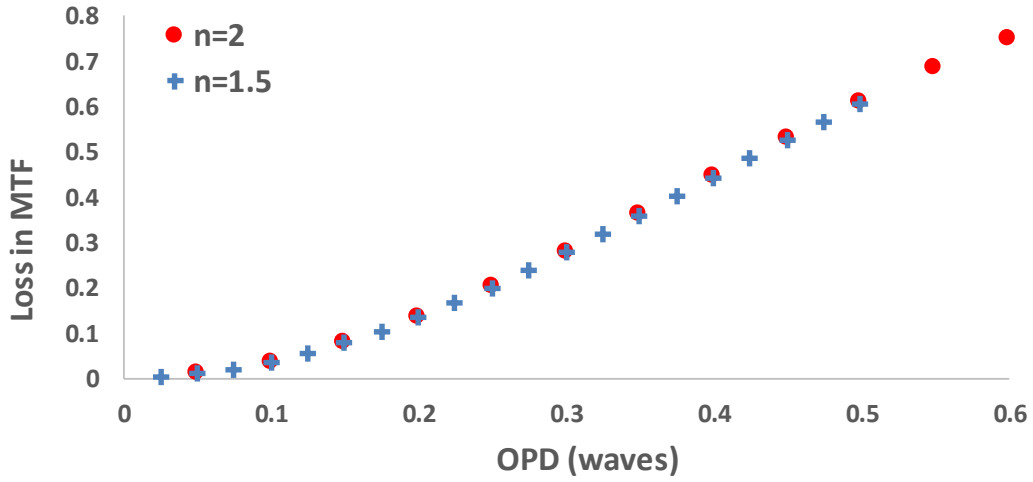


Figure 4.15 The MTF loss *vs.* OPD is the same for surfaces with different refractive indices.

In the next chapter, we will show how Zernike polynomials can be used in predicting optical performance all the way from form to mid-spatial frequency errors.

function

CHAPTER 5 : USING ZERNIKE POLYNOMIALS FOR OPTICAL PERFORMANCE SPECIFICATION

The ability to generate very large orders of Zernike polynomials enables fitting and describing optical surfaces all the way from low order form errors to mid-spatial frequencies. Unlike the power spectral density (PSD), the use of Zernike polynomials does not require windowing and detrending of the surface data for specification and prediction of optical performance.

Our goal is to have a surface characterization metric that fabricators can use to fix their process and designers can incorporate into tolerancing. In the previous chapter, we saw a few examples of typical MSF texture and their effect on modulation transfer function and Strehl ratio. In this chapter, we show how a Zernike polynomial representation of the surface can be used for these calculations. The first section shows the filtering aspect of Zernike polynomials and how different signatures affect optical performance differently. In the second section, we will investigate Strehl ratio as a performance metric for mid-spatial frequency. We will look at Strehl ratios of selected Zernike-filtered bands of a surface and their relation to the Strehl ratio of the entire surface. We will discuss limits of validity of this formulation and present some examples. In the third section, the linear system theory of MTF and limits of its validity for mid-spatial frequency texture is investigated. We will present examples of both real and synthetic MSF textures to examine the approximation limitations of MTF when using linear systems formulation.

5.1 Filtering using Zernike polynomials

Polynomial fitting methods open the door to spatial filtering based on polynomials and their symmetry properties, thus allowing for the isolation of specific MSF structures with different orientations and symmetries caused by different processes. For example, rotationally invariant mid-spatial frequency errors from turning operations are often caused by slowly varying environmental temperature fluctuations such as chiller temperature drift and the impact on tool-tip expansion. Raster patterns in milling operations are usually caused by the overlap in the programmed path of the polishing tool and the tool geometry.

Figure 5.1 (a) shows the height error map of the IRG26 diamond milled sphere, and Figure 5.1 (b) shows the fit residuals after removing the rotationally invariant terms (all ring-like structures present in the map). A clear raster-like pattern is observable. Figure 5.1 (c) shows the rotationally invariant content that was removed. We have used 200 orders of Zernike polynomials to separate these signatures of the surface. Figure 5.2 shows the MTF of the surfaces in Figure 5.1. Knowledge of the optical performance degradation due to different patterns gives designers insight into the ways in which the manufacturing details may give rise to a degradation in optical performance. It also helps fabricators know what parameter of the process needs to be changed.

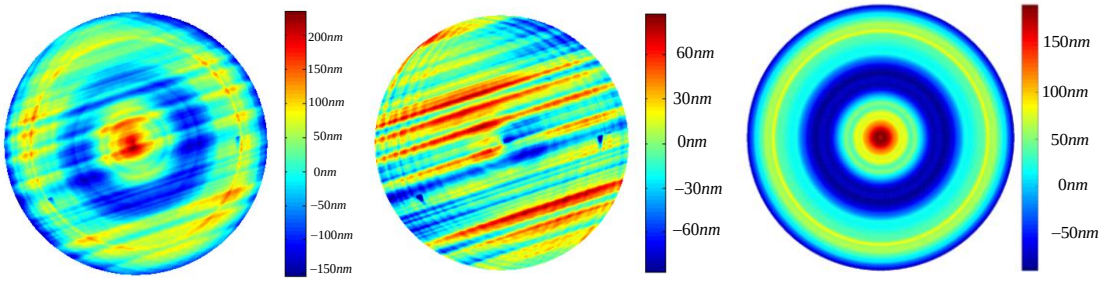


Figure 5.1 (a) Error map of IRG26 diamond milled sphere (b) Error map excluding all rotationally invariant terms (c) Rotationally invariant content of IRG26.

In Figure 5.2 you can see the MTF drop due to rotationally invariant terms and raster patterns of IRG26 sphere separately.

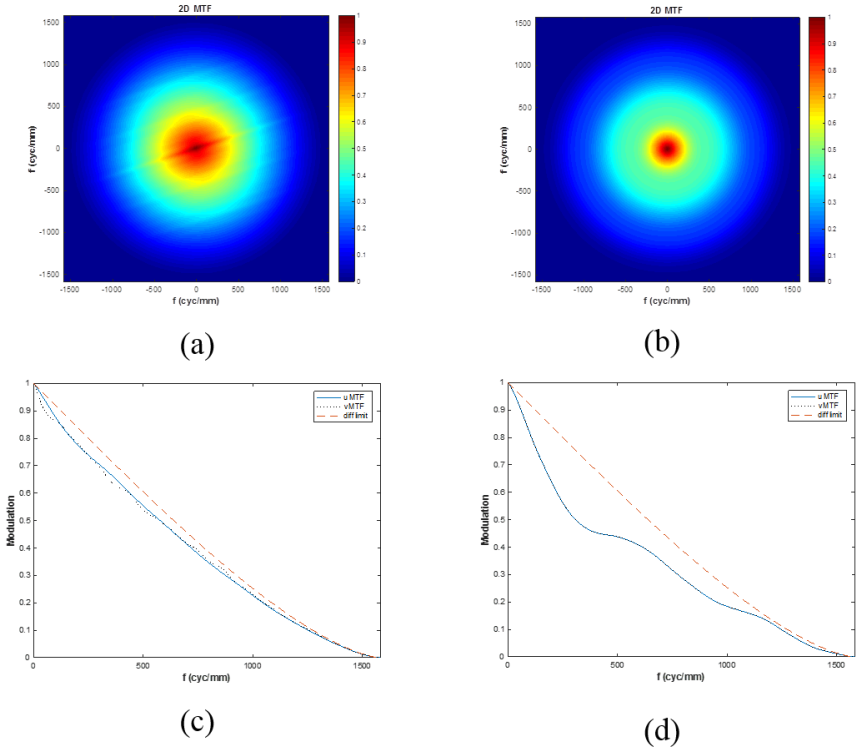


Figure 5.2 (a) The 2D MTF of non-rotationally invariant terms in the IRG26 sphere. (b) The 2D MTF of rotationally invariant terms (c)The cross-sectional MTFs. Vertical and horizontal MTF are not the same for this surface and that is an indication of non-rotationally invariant terms (d) The MTF cross-sections (uMTF and vMTF) for the rotationally invariant content of the IRG26 sphere.

In the simulation for Figure 5.2, we have assumed a reflective surface and a wavelength of 633 nm. The Strehl ratio for the non-rotationally invariant terms is 0.93, and it is 0.77 for the rotationally invariant terms. Therefore, removal of rotationally invariant contributions to the height error map should be of higher priority in improving the fabrication process for this surface. The overall loss is more for the rotationally invariant part, and this contribution to the performance does not depend on the orientation of the part or object. In the presence of raster-like texture on the surface, the performance and the

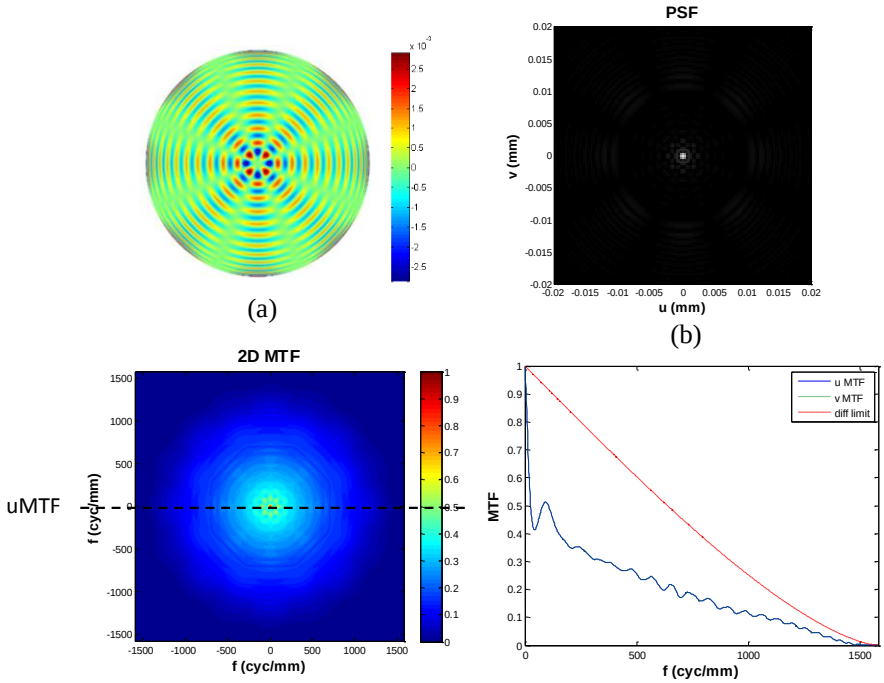
quality of the image varies with the orientation of the surface in the optical system.

5.2 Strehl ratio for mid-spatial frequency errors and limits of validity

The low order Zernike polynomials and their impact on the optical transfer function (OTF) have been discussed in the literature [117]. The effect of large-order Zernike polynomials has not been investigated to our knowledge. In this section, we use the Zernike polynomial characterization of a surface out to high orders to analyze optical performance. V. Mahajan has shown that $e^{-k^2\sigma^2}$ where σ is the RMS of the surface, provides a better approximation for Strehl ratio than the Marechal formula [111]. The Marechal formula approximates the Strehl ratio as $S \simeq (1-k^2\sigma^2/2)^2$. Except for small Strehl ratios, the Marechal formula underestimates the Strehl ratio value. Mahajan has also shown that the error in the Marechal formula is less than 10% if the Strehl ratio is greater than 0.6. The Strehl ratio approximation, $e^{-k^2\sigma^2}$, provides a good estimate with less than 10% error as long as $S \geq 0.3$ [118].

To begin the investigation of extending the Strehl ratio into the MSF range, we look at the PSF, MTF and Strehl ratio for a single high-order Zernike polynomial by adding that this as high order wavefront error to an ideal spherical wavefront. Figure 5.3 (a) shows the error map of the high-order Zernike polynomial Z_{50}^4 with an RMS coefficient of 0.3 waves. Figures 5.3 (b) and (c) show the PSF and 2D MTF of the wavefront shown in part (a). The PSF shows the image of a point source in the presence of this large order aberration. The 2D MTF in (c) shows the image degradation as a function of spatial frequency when the wavefront is aberrated by this very large order Zernike polynomial. In Figure 5.3 (d) we show a horizontal (uMTF) and vertical (vMTF) cross-section of the 2D MTF in part (c). This is an example of a wavefront error where looking at only the horizontal and vertical

cross sections of the 2D MTF is not sufficient to characterize image quality. Due to the symmetry properties of Zernike polynomials, the diversity in MTF with orientation can be assessed with a horizontal cross section, a cross section at an angle of $\frac{\pi}{m}$ and a cross section at $\frac{\pi}{2m}$ from the 2D MTF curve. Figure 5.4 shows the PSF and 2D MTF of a low order Zernike polynomial, Z_3^1 (coma), with the same RMS as Figure 5.3. Comparing the 2D MTF of the high-order Zernike polynomial in Figure 5.3 to coma emphasizes the large number of oscillations in the cross-sectional MTF in Figure 5.3 that is due to oscillations of the wavefront for a high-order polynomial. The Strehl ratio obtained through FFT methods for this aberration, assuming a transmissive surface and a refractive index of 1.51 is 0.42. Calculating the Strehl ratio using Eqn. 4.9 provides the same value.



(a) (b) (c) (d)

Figure 5.3 (a) Large order Zernike polynomial map of Z_{50}^4 (b) Point spread function assuming a transmissive wavefront with a refractive index of 1.51. (c) 2D MTF of the system with Z_{50}^4 aberration on the wavefront (a). (d) Vertical and horizontal MTF of the system.

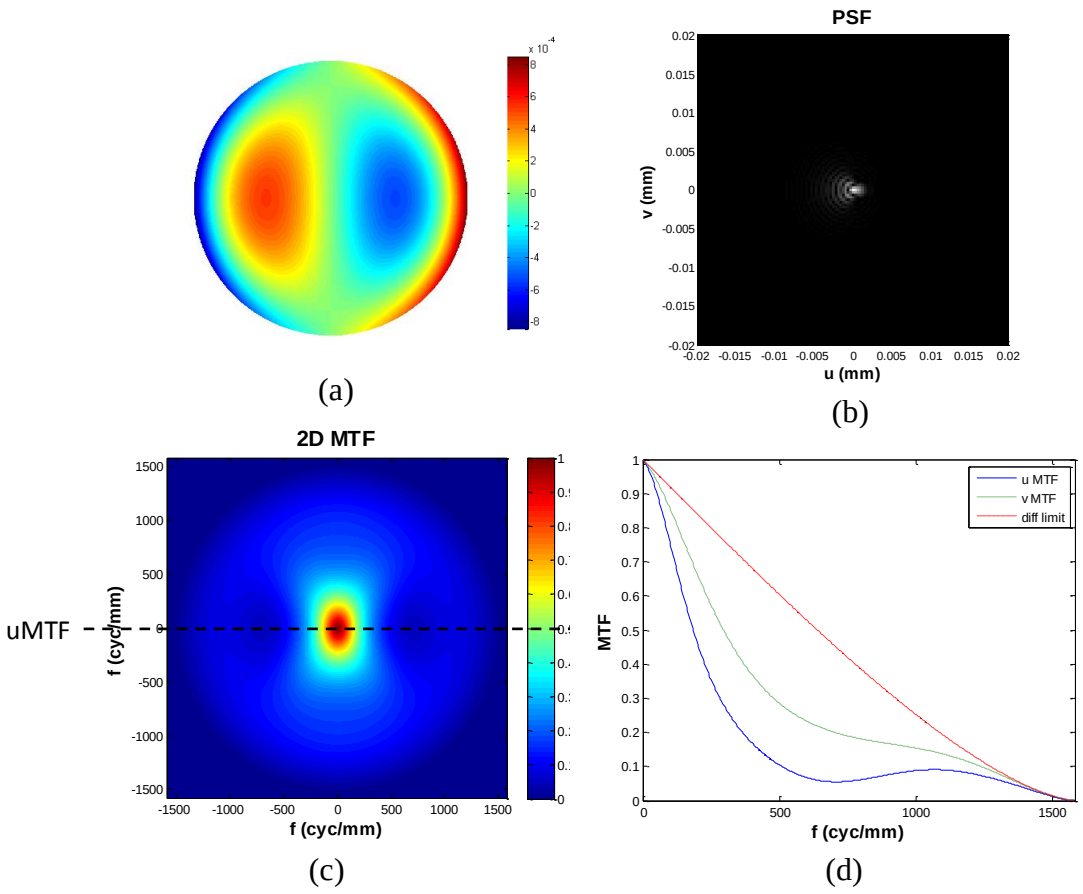


Figure 5.4 (a) Low order Zernike polynomial map of Z_3^1 (b) Point spread function assuming a transmissive wavefront with a refractive index of 1.51. (c) 2D MTF of the system with Z_3^1 aberration on the wavefront (a). (d) Vertical and horizontal MTF of the system.

To verify the validity of Eqn. 4.9 for large orders Zernike polynomials, we have calculated the Strehl ratio for the large order Zernike polynomial Z_{50}^4,using both Eqns. 4.6 and 4.9 as a function of RMS amplitude. Eqn. 4.9 is a convenient calculation for the Strehl ratio, but it is only valid for lower RMS wavefront errors. To quantify its applicability range, we calculate the Strehl ratio directly using FFT methods (Eqns. 4.6 and 4.7) and also using the approximation based on Zernike coefficients ($e^{-k^2\sigma^2}$) of the wavefront Z_{50}^4 as a function of RMS. Figure 5.5 compares the two calculations. As can be concluded from Figure 5.5, the Strehl ratio calculated directly from FFT methods and predicted from Eqn.

4.9 are identical for Strehl ratios as small as 0.1. The benefit of using Zernike polynomial fitting here is that σ^2 (the wavefront RMS) is equal to the root sum square of the fit coefficients when the Zernike polynomials are properly normalized.

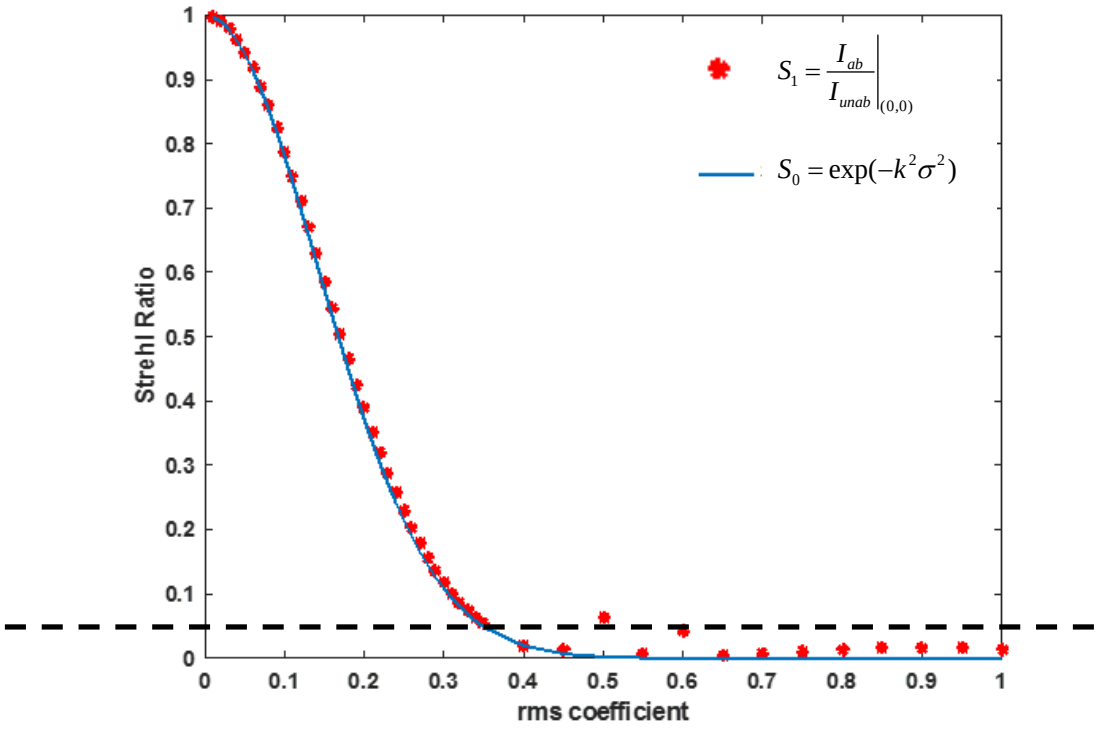


Figure 5.5 Comparison of Strehl ratio calculated directly from Eqn. 4.6 and Strehl ratio predicted based on wavefront variance σ^2 (Eqn. 4.9). Results from two different methods match for Strehl ratios larger than 0.1.

One of the advantages of a polynomial analysis such as Zernike polynomials is the ability to decompose the surface into bands of different texture based on the symmetry properties of the polynomials. Potential error sources during the manufacturing process have varying symmetry characteristics, and these symmetries are transferred to the surface. Figure 5.6 (a-c), shows bands of the aluminum diamond part (Figure 3.9 (a)) along with the 2D MTF maps and their superposition in (d). By using the approximation of Eqn. 4.9 and the orthogonality property of Zernike polynomials, it is straightforward to see that the Strehl

ratio of the sum wavefront is equal to the product of Strehl ratios of each band, as shown in Eqn. 5.1:

$$\begin{aligned} S_w &= \exp(-k^2\sigma_w^2) = \exp(-k^2(\sigma_{w_1}^2 + \sigma_{w_2}^2 + \sigma_{w_3}^2)) \\ &= \exp(-k^2\sigma_{w_1}^2)\times\exp(-k^2\sigma_{w_2}^2)\times\exp(-k^2\sigma_{w_3}^2) = S_{w_1}\times S_{w_2}\times S_{w_3} \end{aligned} . \tag{5.1}$$

The approximation in Eqn. 4.9 is particularly useful when used along with bands of Zernike polynomials since we can benefit from the fact that the sum in quadrature of the Zernike fit coefficients is equal to the RMS of the surface. For example, Figure 5.6 shows azimuthal bands of the aluminum diamond turned part (first row) and their 2D MTF in the second row. The Strehl ratio of the sum wavefront in (d) is equal to the product of the Strehl ratios of the bands in (a), (b) and (c).

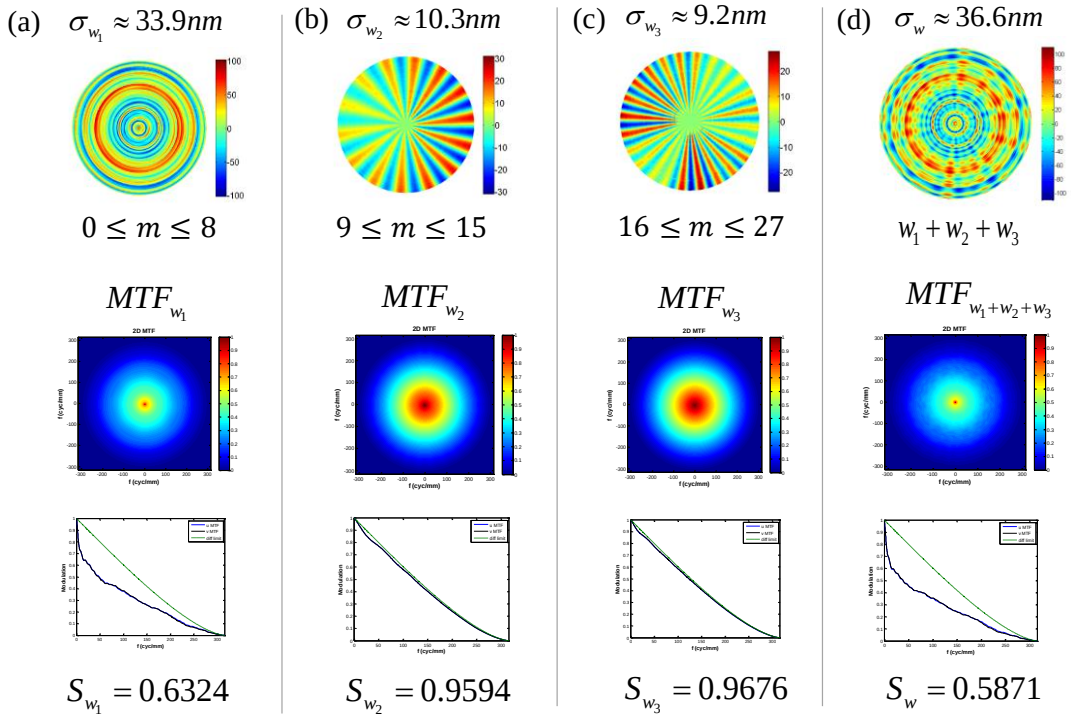


Figure 5.6 Selected bands of aluminum diamond turned part with their 2D MTF and cross-sectional MTFs. (a) Azimuthal band of $0 \le m \le 8$. (b) Azimuthal band of $9 \le m \le 15$ (c) Azimuthal band of $16 \le m \le 27$. (d) Superposition of all three bands. The Strehl ratio of the sum wavefront in (d) is equal to $S_{w_1} \times S_{w_2} \times S_{w_3}$.

When using an orthogonal representation, the surface can also be decomposed into its individual Zernike components such that $W = \sum_{n_z}\sum_{m} \sigma_{n_z}^{\pm m} Z_{n_z}^{\pm m}(\rho,\theta)$. We use a 'Strehl ratio map' to test the validity of the $\exp(-k^2\sigma^2)$ approximation for all Zernike polynomials of a real surface. In Figure 5.7, the Strehl ratios associated with the fit coefficients of SiC sphere (the error map in Figure 3.11 (a)) assuming a reflective surface and incident wavelength of 0.633 μm have been calculated. We have used direct FFT methods for part (a) and indirect methods based on Eqn. 4.9 for part (b). If we construct the SiC sphere based on its 200 Zernike polynomials, the Strehl ratio for that surface is 0.6042. The predicted Strehl ratio based on SiC sphere fit coefficients, i.e. $\prod_{i=1}^{N} \exp(k^2\sigma_i^2)$ is 0.6034. Figure 5.6 (c) shows the difference in Strehl ratios calculated from Eqn. 4.6 and 4.9. The largest difference is smaller than 0.01. The Strehl ratio of the whole surface is 0.6048. This confirms that the error in the approximation to Strehl ratio, $\exp(-k^2\sigma^2)$, is very small. The larger coefficients cause larger loss and that is expected. The Strehl ratio map has the same pattern and trend as a coefficient map and we can predict the Strehl ratio or loss of each band from a fit coefficient map.

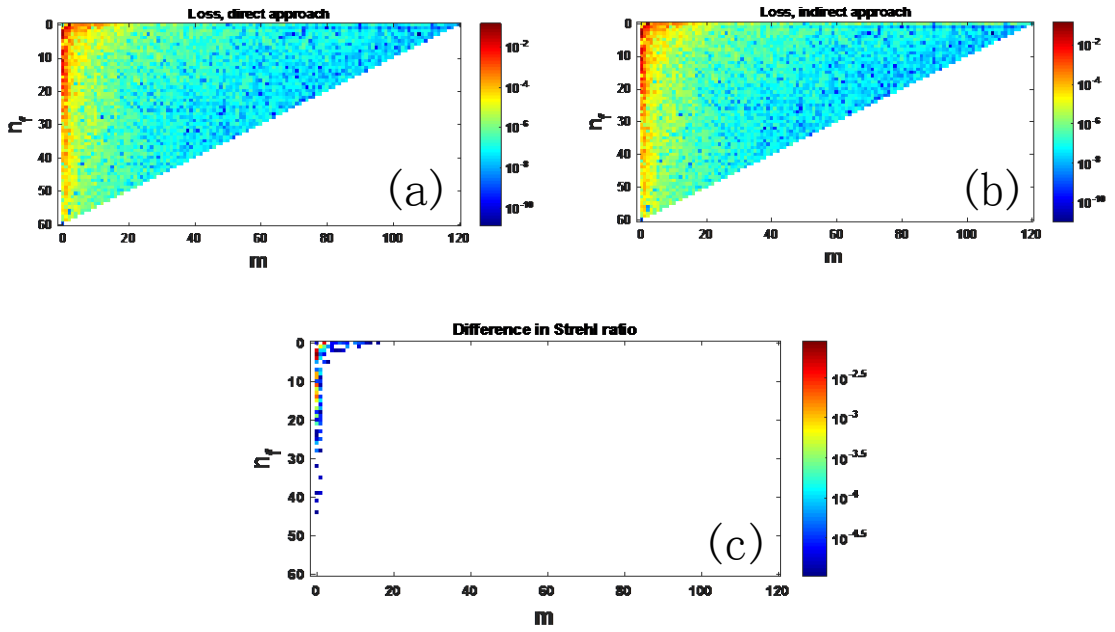


Figure 5.7 (a) The Strehl ratio plot of SiC sphere calculated directly from each coefficient and its corresponding Zernike polynomial map (Eqn. 4.6). (b) Strehl ratio based on a prediction of fit coefficients (Eqn. 4.9) (c) Difference between Strehl ratios calculated directly using FFT methods and based on the prediction.

The Strehl ratio of the sum wavefront is not always equal to the product of Strehl ratios of its components. An example of this is shown in Figure 5.8. Figure 5.8 (a) is the first 36 Zernike terms of diamond milled germanium sample in Figure 3.13 (a). Figure 5.8 (b) is the fit residual errors of the germanium surface when subtracting 36 terms of Zernike polynomials. The Strehl ratio calculated for the whole surface is different from the product of the Strehl ratios in part (a) and part (b). In this case, the Strehl ratio of one of the bands (W_1) is 0.069, which is below 0.1 and therefore we do not expect the approximation to the Strehl ratio $e^{-k^2\sigma^2}$ to be valid, consequently $S_{\sum_i w_i} \neq \prod_i S_{w_i}$. The row below the surface profiles in Figure 5.8 depicts cross sections of the 2D MTF. As can be seen, the drop in MTF in Figure 5.8 (a) is very large and the MTF is close to zero within most of the frequency range. The large drop in the MTF of the wavefront in (a) is consistent with a

breakdown in the approximation to the Strehl ratio, and as a result, the Strehl ratio of the sum wavefront (0.073) is not equal to the product of Strehl ratios of its components (0.06). The simulations for this figure are carried out using the Mx software and the assumptions are described in section 4.1.

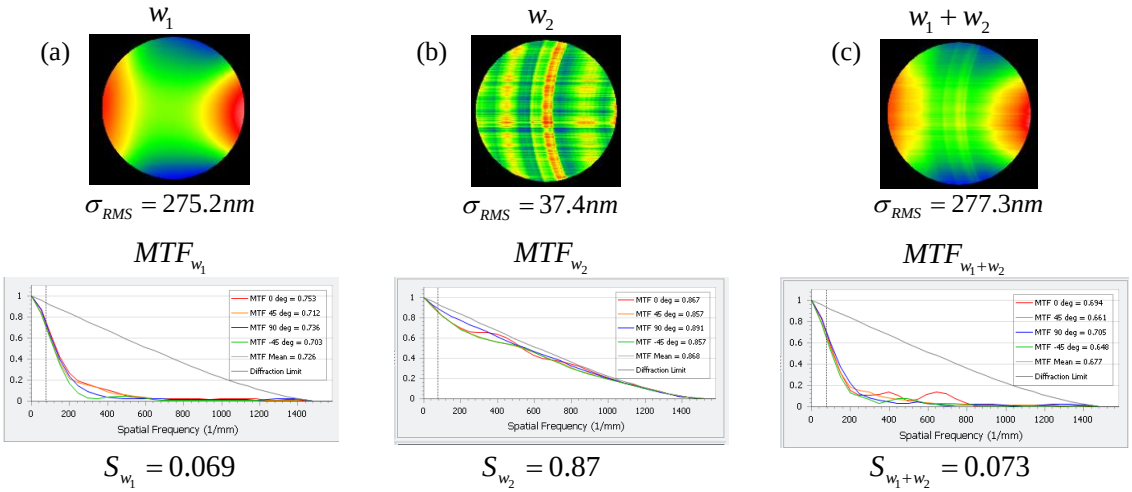


Figure 5.8 Selected bands of diamond milled germanium part. The second row is cross-sectional MTF of each band. (a) The first 36 Zernike terms of the sample in Figure 3.13 (a). (b) The fit residual errors after subtracting 36 terms of Zernike polynomials. (c) The superposition map and its MTF and PSF. $S_{w_1} \times S_{w_2} = 0.06$ and is not equal to 0.073.

The main advantage of decomposing the surface into bands of polynomials is that the metrologist can know what aberrations with what magnitude are present in the system and how much each band affects the optical performance. Furthermore, the Strehl ratio approximation with its direct relation to surface RMS provides a means of assessing the optical performance with only an analysis of the surface without needing to perform a Fourier-based calculation. The user must keep in mind that this prediction is not valid if the coefficients (surface errors) are large enough to give rise to a Strehl ratio that is smaller than 0.1. Because the Strehl ratio depends only on wavefront RMS, two wavefronts can have the same RMS (and therefore Strehl ratio) but the spatial frequency content of the wavefront can be very different. This means the point spread functions can be very

different. Thus, the use of the Strehl ratio of the full surface error is not particularly useful when MSF texture is present. Strehl ratio as an MSF specification metric might be reasonable, however, if the wavefront is decomposed or filtered into the low order aberration terms and the MSF errors, separately. Then, separate Strehl ratio specifications could be sensibly defined for each band.

This problem of using the full wavefront RMS (or Strehl ratio) alone has been discussed by J. Tamkin [24]. J. Tamkin has shown that 0.66 waves defocus, causes the MTF loss of 20%. Mid-spatial frequency errors with the same RMS will result in 30% loss in MTF and a cyclic MTF curve [24]. He suggests using characteristics of the MTF curve for specification, such as specifying an allowable drop in the MTF for MSF errors on optical surfaces [24]. MTF provides more detailed information on the frequency response of an imaging system and is one of the most common methods to specify optical elements [114]. Since MSF errors are usually quasi-periodic with specific symmetry and anisotropy characteristics, considering the impact of these errors in terms of the frequency response seems reasonable. In the next section, we will see examples of the MTF response of the system in the presence of mid-spatial frequency errors, and we will investigate the limitations of linear systems theory of MTF and its application to MSF texture.

5.3 Linearity properties of MTF and application to mid-spatial frequency error

According to linear systems theory, if components that lead to image degradation are independent and uncorrelated, the system's OTF is given by the product of the OTFs of individual components [114, 119]. If the sub-components in the system respond to the image degradation without any partial coherence effects, the MTF of the aggregate system is the point-by-point product of individual MTFs [114]. The sub-components will not

change the partial coherence of the intermediate image if their superposition gives no visible fringes [120], and that is the case if sub-components are not correlated [48]. That is, if the mutual coherence function between the wavefronts of the sub-components is zero, we can multiply their MTFs to obtain the MTF of the aggregate system [121]. By normalizing the mutual coherence function (Eqn. 5.2) to the product of the square root of the intensities, we obtain the correlation coefficient function as shown in Eqn. 5.3 [48].

$$\Gamma_{12}(\tau) = \left\langle V_1(t+\tau)V_2^*(t) \right\rangle \tag{5.2}$$

$$r_{12} = \frac{\Gamma_{12}}{\sqrt{\Gamma_{11}}\sqrt{\Gamma_{22}}} = \frac{\Gamma_{12}}{\sqrt{I_1}\sqrt{I_2}}. \tag{5.3}$$

If the above assumptions are valid, the convolution theorem tells us that the system's PSF is equal to the convolution of the PSFs of individual components,

$$\mathbb{H}_{system} = \mathbb{H}_{w_1} \times \mathbb{H}_{w_2} \times \mathbb{H}_{w_3} \times \mathbb{H}_{w_4} \text{ and} \tag{5.4}$$

$$PSF_{system} = PSF_{w_1} \otimes PSF_{w_2} \otimes PSF_{w_3} \otimes PSF_{w_4}. \tag{5.5}$$

J. E. Harvey *et al*. have developed a nonparaxial scalar diffraction theory to develop a linear systems formulation of surface scatter phenomena [122]. Unlike previous methods such as Rayleigh–Rice or classical Beckmann–Kirchhoff, the generalized Harvey Shack model does not have the smooth and isotropic surface assumptions [123]. Without the smooth and isotropic assumptions, the linear systems formulation can be applied to MSF texture.

In this section, we investigate the MTF of filtered bands in measured surfaces to test the linearity property of MTF for isolated textures. Figure 5.9 shows two filtered bands of the germanium sample in Figure 3.13 (a), part (a) Cartesian bands of $t \le 13$ with

$\sigma_{RMS} \approx 270nm$ and part (b) Cartesian bands of $25 \le t \le 30$ with $\sigma_{RMS} \approx 10.9nm$. The second row in Figure 5.9 shows the 2D MTF of each surface and the third row shows horizontal and vertical cross sections of the 2D MTF.

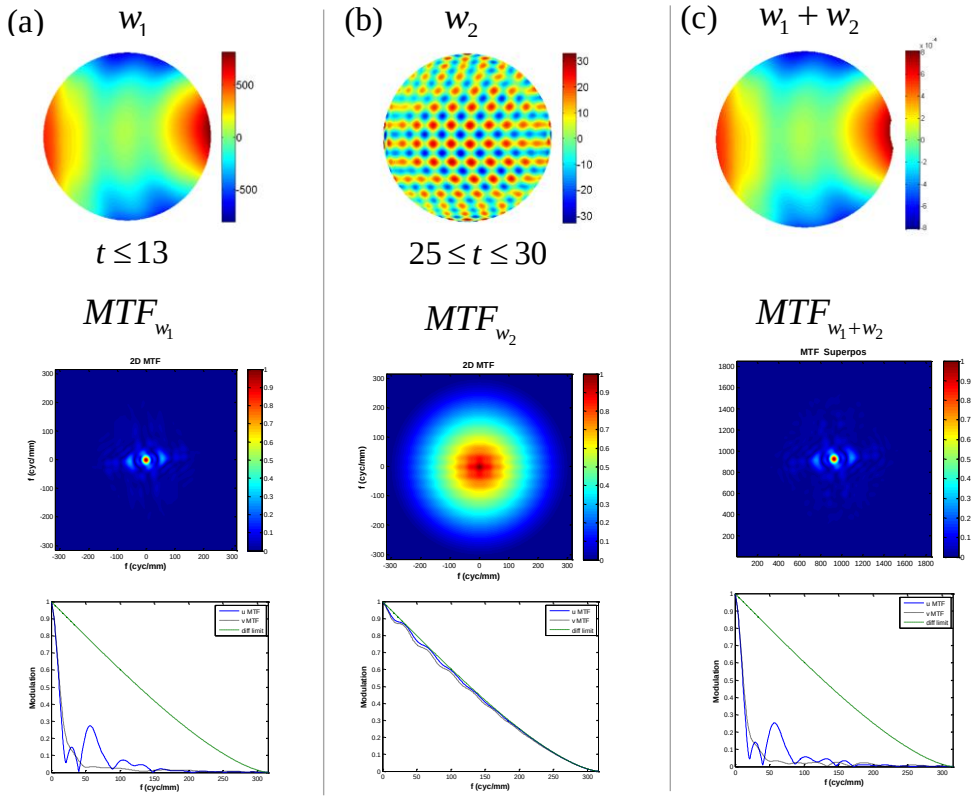


Figure 5.9 Filtered bands of germanium sample, their superposition and their MTFs. (a) Cartesian bands of $t \le 13$ with $\sigma_{RMS} \approx 270nm$ with its MTF depicted below the surface (b) Cartesian bands of $25 \le t \le 30$ with $\sigma_{RMS} \approx 10.9nm$ with its MTF and cross-sectional MTF shown below the surface error. (c) Superposition surface map and its MTF.

Eqn. 5.2 and Eqn. 5.3 show that if the cross-correlation between two wavefronts is zero, we can apply linear system theory of MTF. The cross-correlation coefficient between the two wavefronts in Figure 5.9 (a) and (b) is -0.0026 and is obtained using Eqn. 5.6.

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2\right)\left(\sum_m \sum_n (B_{mn} - \bar{B})^2\right)}}, \tag{5.6}$$

assuming A and B are matrices that represent W_1 and W_2. Eqn. 5.6 shows another definition of cross-correlation that is the equivalent of Eqn. 5.3. Since the correlation coefficient between these two surfaces is negligible, we examine to see if the linear systems formulation of MTF is valid in this case. Simply multiplying MTF_{w_1} and MTF_{w_2}, would count the diffraction limited MTF twice since aperture diffraction is present in every band [114]. Therefore, to carry out this comparison, the diffraction limited contribution is accounted for in MTF_{1+2} and the error $\varepsilon = \left[MTF_1 \times MTF_2\right] - \left[MTF_{1+2} \times MTF_{diff}\right]$ is plotted and shown in Figure 5.10. The error in the linear systems formulation of MTF is negligible in this example where the correlation between two surfaces is small (-0.0026).

$$\left[MTF_1 \times MTF_2\right] - \left[MTF_{1+2} \times MTF_{diff}\right]$$

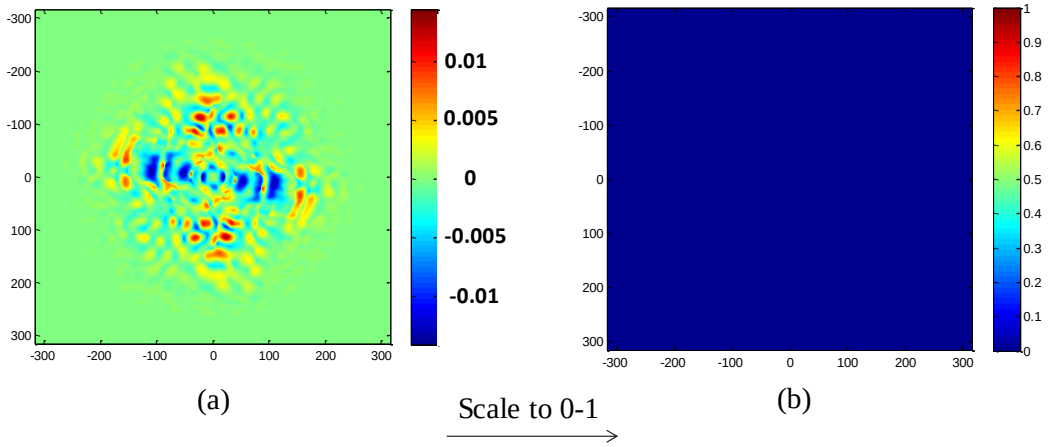


Figure 5.10 (a) The difference between the product of MTFs of wavefronts in Figure 5.8 (a) and (b), and the MTF of their superposition. (b) The error is plotted here on the same scale as an MTF plot (from zero to one) and it is clearly negligible.

To find the geometrical MTF of any system or subsystem, the aperture diffraction contribution should be separated [114]. To obtain the geometrical MTF, we should normalize the MTF of the aberrated wavefront by its diffraction-free MTF as shown in Eqn. 5.7.

$$MTF_{w,geom} = \frac{MTF_w}{MTF_{w,diff}}. \tag{5.7}$$

Figure 5.11 shows the geometrical MTF of different bands of the germanium sphere. Figure 5.11 (a) shows the wavefront that contains the band of $25 \le t \le 30$ along with its MTF in the second row and its geometrical MTF in the third row. Figure 5.11 (b) shows a different selected band of the germanium sample, $35 \le t \le 45$. Part (c) shows the superposition of the wavefronts in (a) and (b). In Figure 5.11 (d), one can see that the product of the geometrical MTFs of the two bands overlays the geometrical MTF of the sum wavefront. i.e. $MTF_{1+2,geom} = MTF_{1,geom} \times MTF_{2,geom}$.

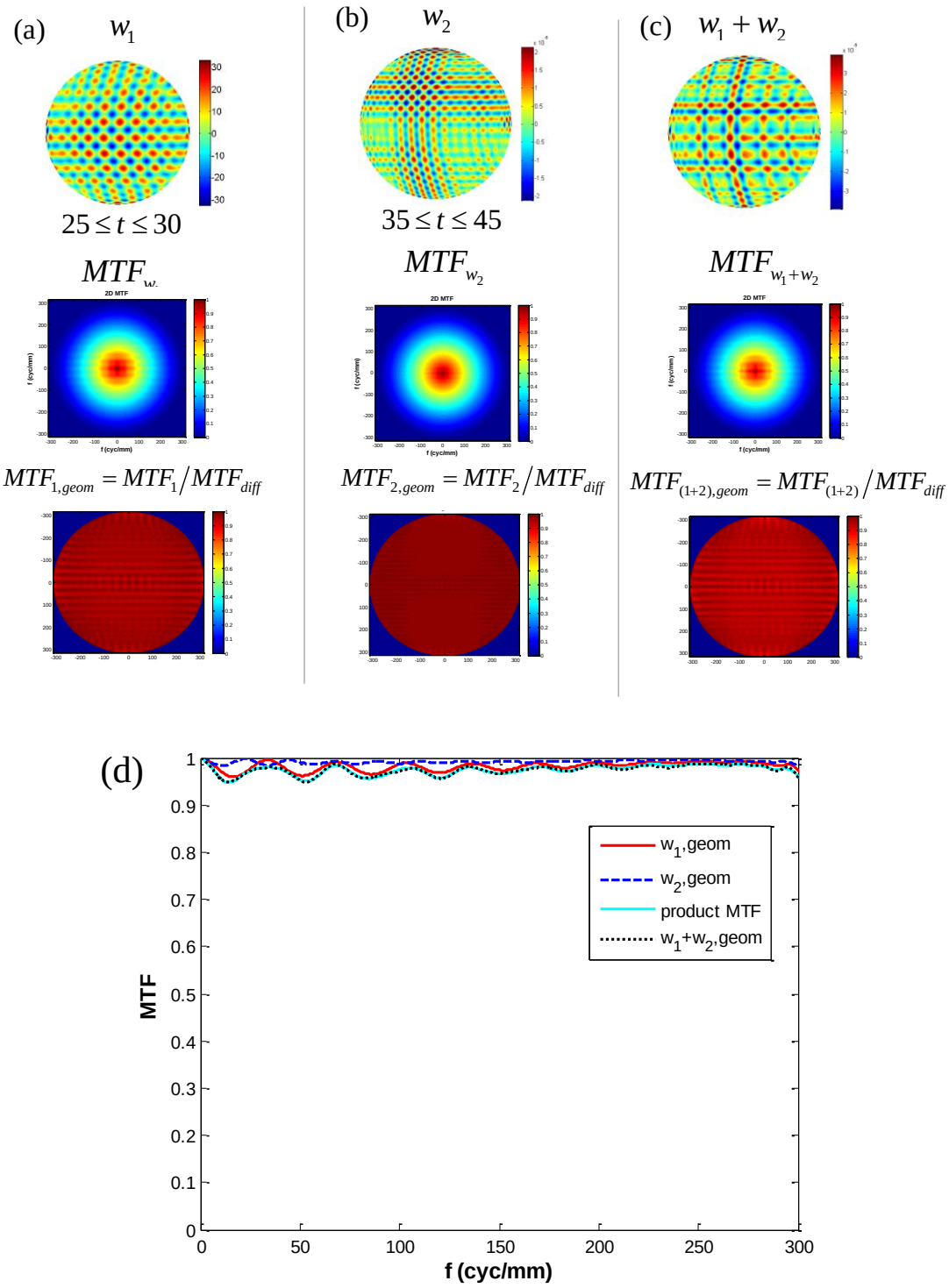


Figure 5.11 (a) Filtered band of the germanium sample $25 \le t \le 30$ with its MTF in the second row and its geometrical MTF in the third row. (b) Filtered band of the germanium sample $35 \le t \le 45$ with its MTF in the second row and its geometrical MTF in the third row. (c) Superposition of the bands in (a) and (b). (d) Cross sections of the geometrical MTFs in a-c and the product of geometrical MTFs of (a) and (b).

The linear properties of MTF are only applicable when the correlation between the components of the system is negligible. In an attempt to investigate the validity of linear systems properties of MTF, bands of real surfaces, as well as artificially generated bands and their superposition, have been examined. We take the error in the MTF calculation using linear systems formulation approximation to be the difference between the correct value (MTF of sum wavefront) and the approximated value (the product of individual MTFs), normalized by the correct value:

$$\varepsilon_{MTF} = \frac{\iint_{f_u, f_v} \left| \left[MTF_1 \times MTF_2 / MTF_{diff} \right] - \left[MTF_{1+2} \right] \right| dudv}{\iint_{f_u, f_v} \left[MTF_{1+2} \right] dudv} . \tag{5.8}$$

We applied this calculation to a variety of filtered real surface data and simulated data, and the findings are plotted in Figure 5.12 where the error is plotted *vs.* the correlation coefficient between the individual bands. Table 5.1 provides more information about each data point in the plot.

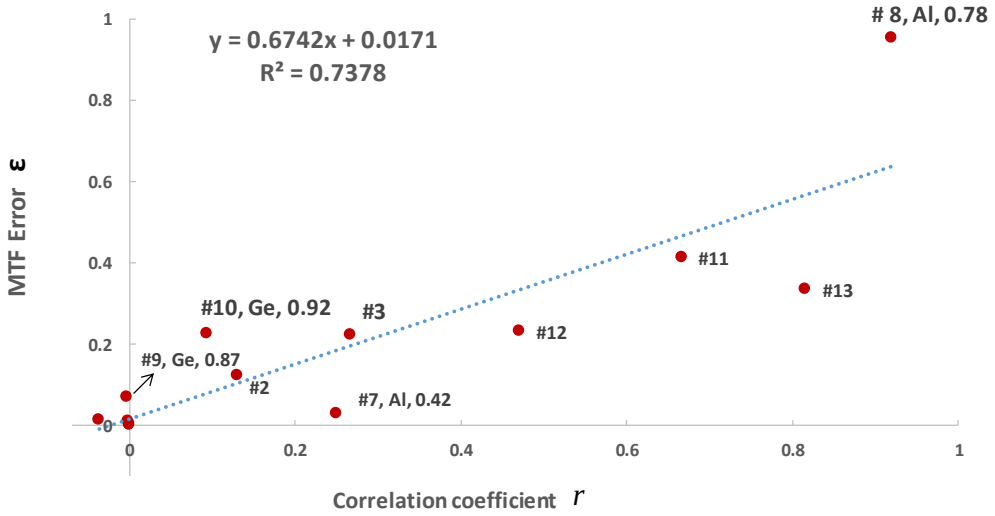


Figure 5.12 (a) The error is MTF approximation obtained from Eqn. 5.8 plotted against the correlation coefficient between individual surfaces composing the system. The raw data are presented in Table 5.1.

Each data point in Figure 5.12 and Table 5.1 is obtained from the MTF of the surfaces shown in Figures 5.13 and 5.14.

Table 5.1 Data obtained from 13 different cases in Figure 5.12. The surfaces are chosen from selected bands of real surfaces as well as artificially generated bands.

Sample #	Correlation coefficient r	MTF Error $\mathcal{E}$	Max MTF drop	Average MTF drop
1	-0.037	0.0144	0.5478	0.0969
2	0.1311	0.1218	0.6056	0.1232
3	0.2675	0.2228	0.4394	0.109
4	-9.10E-04	0.0109	0.144	0.0322
5	3.84E-06	0.0027	0.3414	0.0771
6	-1.12E-05	0.0013	0.0795	0.0141
7	0.25	0.0306	0.4209	0.0918
8	0.92	0.9551	0.7874	0.1724
9	-0.0026	0.0688	0.8723	0.1852
10	0.0938	0.2269	0.9289	0.1841
11	0.6667	0.4125	0.7498	0.1671
12	0.4714	0.2312	0.6953	0.1523
13	0.8165	0.3345	0.7306	0.1576

In Figure 5.13, cases (1) to (3) are raster patterns that are artificially generated using the diagonal bands of Zernike-fit coefficient maps. The surfaces in case (4) are generated using bands of specific azimuthal orders. Cases (5) to (8) are bands of aluminum diamond turned sample in Figure 3.9 (a). The correlation between the wavefronts of the case (8) is 0.9 despite no overlapping bands. This is due to the fact that the fit coefficients of this sample in the region of $0 \le m \le 8$ does not converge to zero and as a result, there will be a significant amount of $0 \le m \le 8$ left in the residual map. The band of $28 \le m$ contains all the fit residuals of 300 orders that include residuals of non-converging bands. That is why the cross-correlation value between the non-overlapping bands of the case (8) is large.

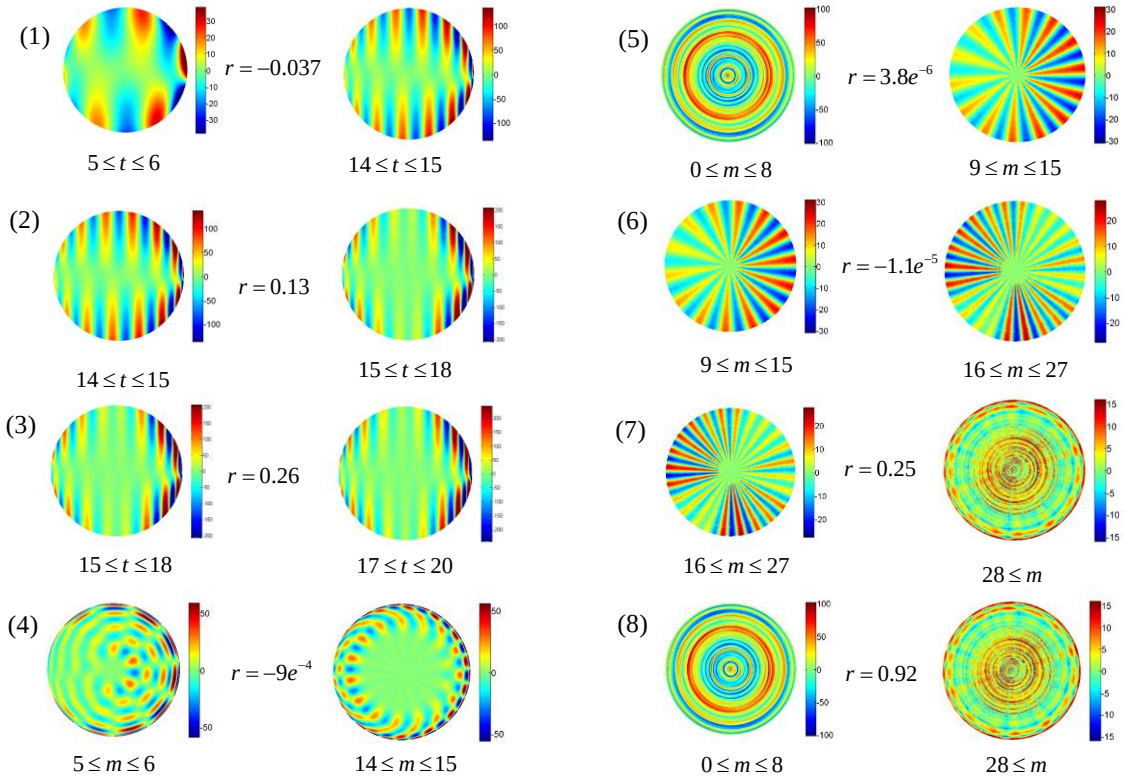


Figure 5.13 (a) Surface error maps with their correlation coefficients. (1-3) Raster patterns generated artificially from bands of different Cartesian orders. (4) Error maps of azimuthal bands (5-8). Selected bands of diamond turned aluminum sample.

In Figure 5.13, cases (9) and (10) are selected bands of germanium sample. Case (10) has some overlapping bands. Cases (11) to (13) are spokes patterns with different relative phases.

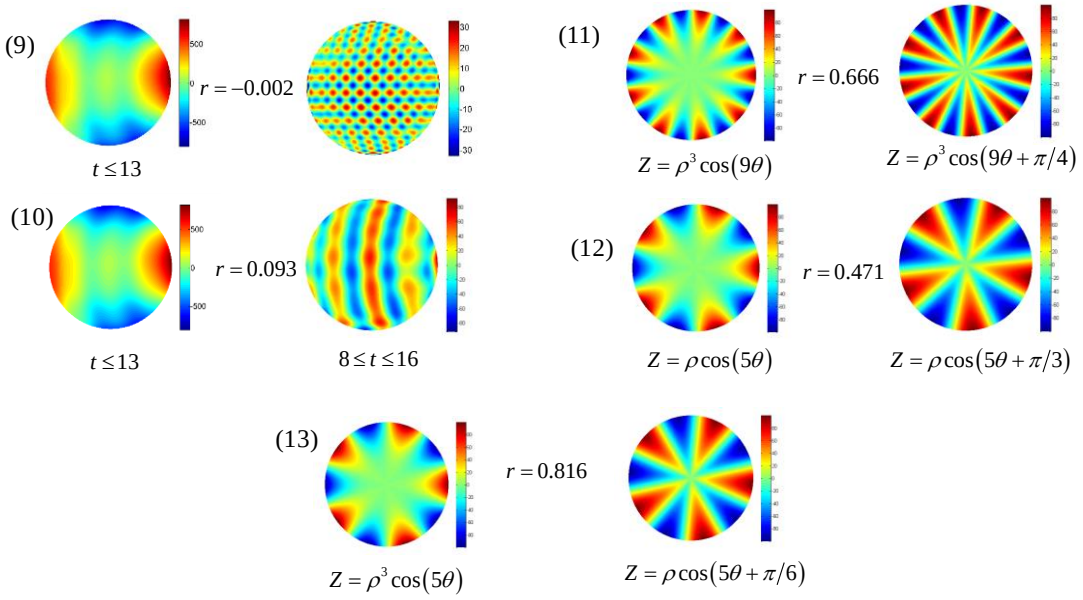


Figure 5.14 Surface error maps of Table 5.1 with their correlation coefficients. (9-10) Cartesian bands of germanium sample. (11-13) Artificially generated spokes patterns with different relative phases.

Figure 5.11 shows that the error in MTF is connected to the correlation between individual surfaces. While there is a connection between correlation and the error in the approximation of MTF using linear systems theory, Eqn. 5.8 does not capture enough detail to provide a clear assessment of the error when wavefronts are partially correlated. Further work could be done to establish a more detailed model for such cases. When surfaces (or wavefronts) are uncorrelated, the error in the validity of linear systems formulation of MTF is small and we can benefit from the cascade properties of MTF. The data points that are above the fitted line in Figure 5.12, for example, cases number 8, 9 and 10 are surfaces where the error in approximation of MTF is more sensitive to the correlation between the components. The surface errors of these cases are shown in Figures 5.13 and 5.14. These

cases have a few characteristics in common. They have form errors in one of their components. They also have larger maximum MTF drop (MTF drop greater than 0.75). Larger MTF drop allows for a larger error in the approximation of MTF when specifying band-limited surfaces.

The orthogonality property of Zernike polynomials provides us with an effective tool to separate bands of a surface with textures caused by different fabrication processes. The linear systems formulation of MTF provides us with another useful tool for specifying tolerable wavefront errors and therefore tolerable surfaces. This tool can be applied to cascaded components of a system as well as uncorrelated bands within an individual sub-system or wavefront. However, orthogonality does not guarantee zero correlation. The correlation coefficient between the two wavefront maps W_1 and W_2 is related to the mean of their inner product and the mean of each map. The inner product of two orthogonal maps is zero.

$$r[w_1, w_2] = \frac{Cov[w_1, w_2]}{\sigma[w_1]\sigma[w_2]} = \frac{\langle w_1.w_2 \rangle - \langle w_1 \rangle \langle w_2 \rangle}{\sigma[w_1]\sigma[w_2]} \tag{5.9}$$

Therefore, for two maps of zero mean, orthogonality results in zero correlation. Since Zernike polynomials are orthogonal, the mean of the dot product of selected bands is zero. The mean of Zernike polynomials is zero for comma, astigmatism and non-rotationally invariant terms [97].

$$\left\langle Z_n^m(\rho,\theta) \right\rangle = \frac{\int_0^1 \int_0^{2\pi} Z_n^m(\rho,\theta)\rho d\rho d\theta}{\int_0^1 \int_0^{2\pi} \rho d\rho d\theta} = 0, \qquad n \neq 0,\ m \neq 0. \tag{5.10}$$

Zero mean plane differs from the reference sphere for rotationally invariant terms and therefore, their mean is not zero. In order to apply the linear systems formulation of MTF along with filtering capability of Zernike polynomials, all of the rotationally invariant components of the surface (wavefront) must be in only one of the bands. As long as one of the bands contains all of the rotationally invariant terms (all *m=0* Zernike terms), the correlation between the bands will be negligible and therefore the linear systems properties of MTF applies.

$$MTF\left[\exp(ikW)\right] = MTF\left[\exp(ik\sum_{n,m}\sigma_n^m Z_n^m)\right] = \prod_{n,m} MTF\left[\exp(ik\sigma_n^m Z_n^m)\right]. \qquad (5.11)$$

Figure 5.15, shows filtered bands of the aluminum diamond-turned part. Figure 5.15 (a) contains all the rotationally invariant terms ($m = 0$) and Figure 5.15 (b) is the azimuthal bands of $9 \le m \le 15$. The mean of the surface in (a) is $-0.73nm$ -a non-zero value. The mean of the surface in (b) is $5.8\times e^{-7} nm$ t- a negligible value. The mean of their dot product, $\langle w_1.w_2 \rangle$, is equal to $2.86\times e^{-5}$. Eqn. 5.12 shows the correlation coefficient between surfaces in Figure 5.15 (a) and (b) that is equal to the correlation coefficient obtained from Eqn. 5.9.

$$r(w_1, w_2) = \frac{2.86\times e^{-5} - (-0.73)\times(5.8\times e^{-7})}{28.3\times 9.05} = 1.13\times e^{-7} \qquad (5.12)$$

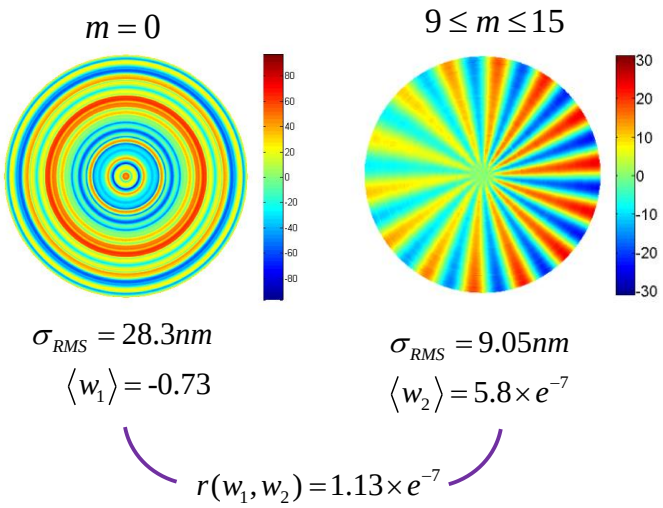


Figure 5.15 (a) Rotationally invariant bands of the aluminum diamond turned part with RMS of 28.3 nm. (b) Azimuthal band of $9 \le m \le 15$ with RMS of about 9 nm.

This example verifies that as long as one of the filtered bands is composed of only non-rotationally invariant terms, then the correlation coefficient between orthogonal bands of Zernike polynomials of the surface is close to zero and we can apply the linear systems formulation of MTF in specifying each band.

5.4 Application to measured data

In this section, we apply the linear systems theory of MTF to filtered bands of measured wavefronts. We will show different filtering examples based on the conclusion in section 5.3. To benefit from the linear properties of MTF, we separate the rotationally invariant content of the surface and apply other filters on the non-rotationally invariant residuals based on the signature of the coefficient map. This will then allow us to specify MTF of each band separately. Figure 5.16 shows different azimuthal bands of the aluminum diamond part sample. Figure 5.16 (a) is all of the rotationally invariant terms plus the azimuthal band of $28 \le m$. The fit coefficients of the error map do not converge to zero even for 300 orders (see Figure 3. 9 (b)). Therefore, there will be significant rotationally invariant content in the residual map of $28 \le m$ and therefore a large correlation between

these two bands (r=0.3). Thus, we have added the residual band of $28 \le m$ to the rotationally invariant terms as shown in Figure 5.16 (a) to ensure all rotational invariant structure is isolated into one band. Parts (b), (c) and (d) show the azimuthal bands of $1 \le m \le 8$, $9 \le m \le 15$, and $16 \le m \le 27$ with the RMS of 10.35nm, 10.32 nm and 9.2 nm respectively. The wavefront in (e) is equal to the superposition of the wavefronts in (a) to (d) as shown in Eqn. 5.13. The MTF of the sum wavefronts in (e) is almost equal to the product of the MTFs of filtered bands in (a) to (d) (Eqn. 5.14).

$$w = \sum_{i=1}^{N} w_i \tag{5.13}$$

$$MTF^{g}_{1+2+...+N} \approx \prod_{i=1}^{N} MTF^{g}_{i} \tag{5.14}$$

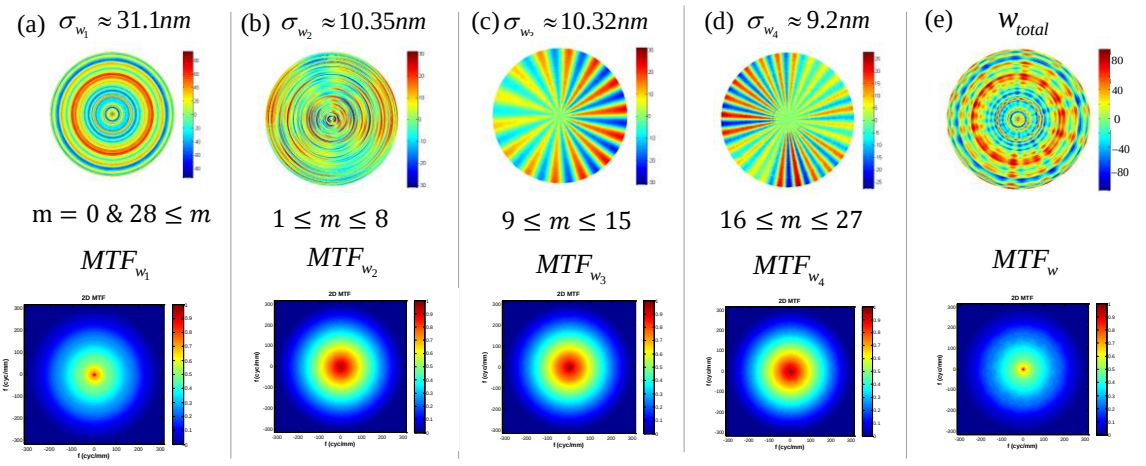


Figure 5.16 Azimuthal bands of the aluminum diamond turned part in the first row and their corresponding MTF in the second row (a) Rotationally invariant bands and the azimuthal band of $28 \le m$ with the RMS of 31.1 nm (b) Azimuthal band of $1 \le m \le 8$ with the RMS of 10.35 nm and its 2D MTF in the second row (c) Azimuthal band of $9 \le m \le 15$ with RMS of about 10.32 nm. (d) azimuthal band of $16 \le m \le 27$ with the RMS of 9.2 nm. (e) Superposition of the wavefronts in (a) to (d).

Figure 5.17 shows the error in the calculation of MTF by obtaining the difference between the MTF of the sum wavefront and the product of MTFs of all the bands. We have

accounted for the diffraction-limited MTF in simulating the MTF error in Figure 5.17 by multiplying the MTF of the sum wavefront by MTF_{diff}^{n-1} where n is the number of filtered bands. The scale in Figure 5.17 (a) is the order of 10^{-3}. Changing the color bar scale to [0 1] in Figure 5.17 (b) shows that the error in the approximation of MTF is negligible. The volume MTF error is obtained through Eqn.5.8 and is equal to ~0.25%.

$$\left[MTF_1 \times MTF_2 \times MTF_3 \times MTF_4\right] - \left[MTF_{1+2+3+4} \times MTF_{diff}^3\right]$$

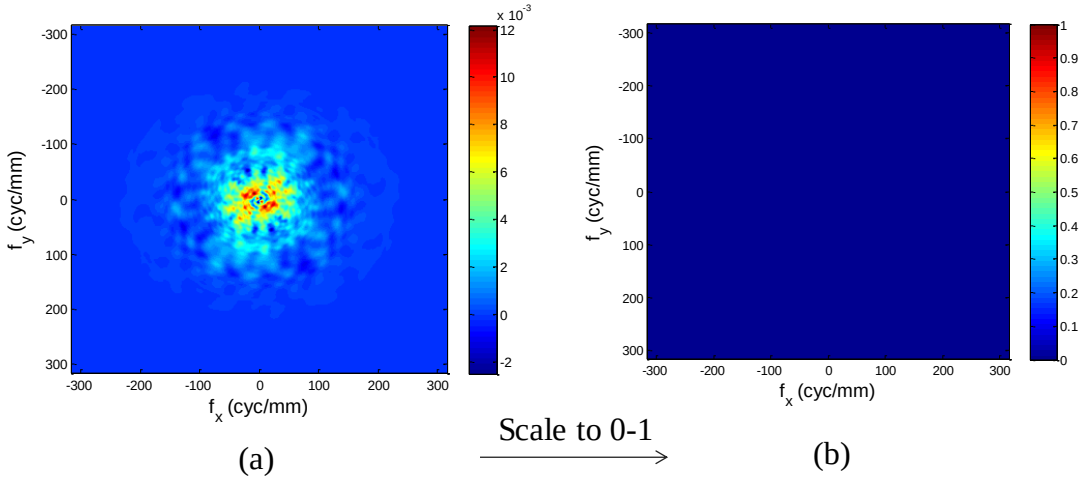


Figure 5.17 (a) Error in the approximation of MTF obtained by subtracting the MTF of the superposition wavefronts in Figure 5.16 and the product of MTFs of all bands. The scale is in the order of 10^{-3}. (b) The MTF error when we change the color scale to [0 1]. The difference is negligible.

The previous example had signatures of spokes-like patterns. To examine the linear system theory of MTF for filtered bands of raster patterns, we have selected different bands of the germanium sample. Prior to filtering of the germanium sample error map, the primary astigmatism terms have been removed due to their large RMS so that MSF textures in different bands and their MTF will be more visible. The filtered band in Figure 5.18 (a) is all of the rotationally invariant terms. Figures 5.18 (b), (c), (d), (e), (f) and (g) show the diagonal bands of $t \le 14$, $15 \le t \le 35$, $36 \le t \le 50$, $51 \le t \le 78$, $79 \le t \le 110$,

$111 \le t \le 119$, and $120 \le t$ with their RMS shown above each map. The filtered bands in (b) to (h) exclude the rotationally invariant terms. Figure 5.17 (i) shows the error map of the germanium sample where the astigmatism terms ($Z_2^{\pm 2}$) have been removed and this is also equal to the superposition of the filtered bands from (a) to (h).

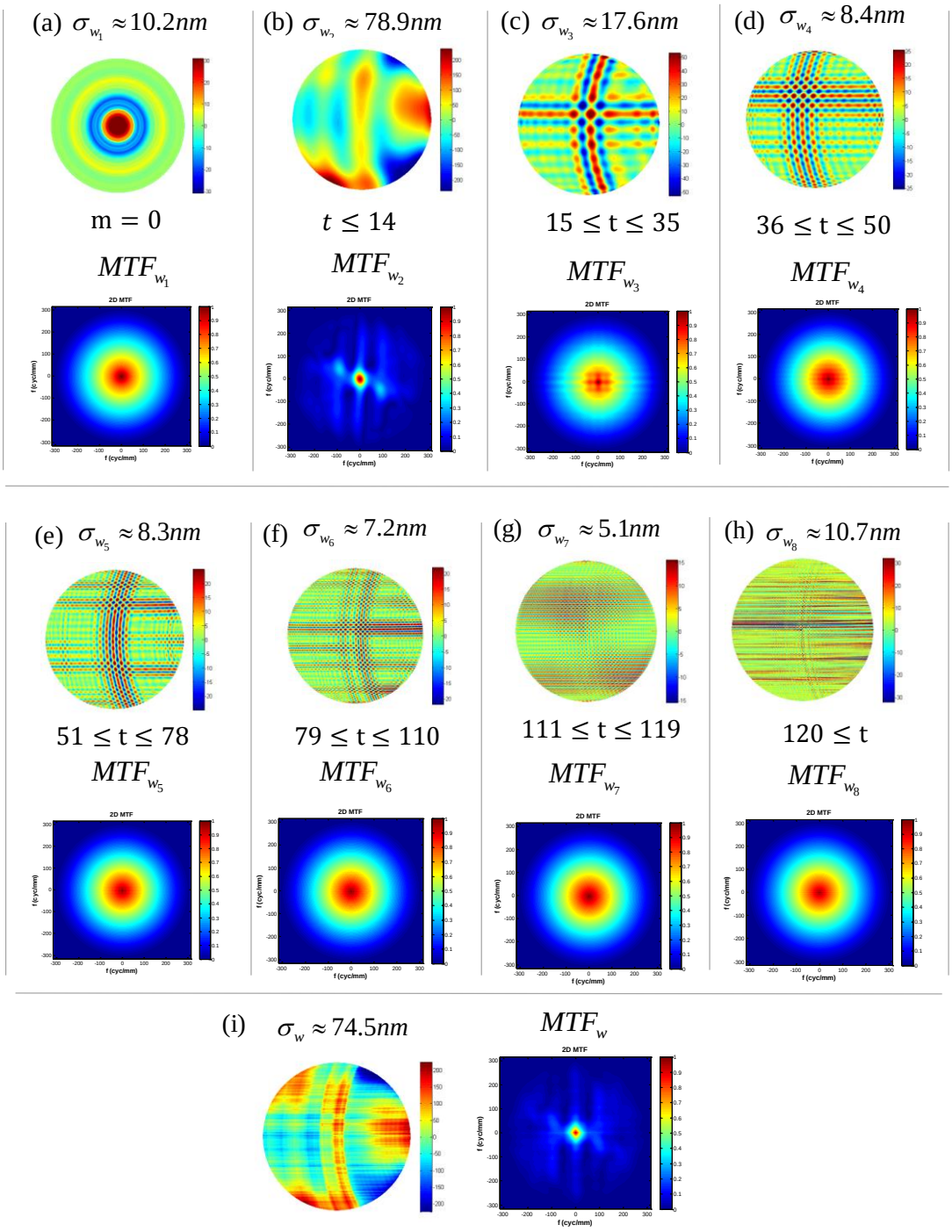


Figure 5.18 Diagonal t-bands and the rotationally invariant content of the germanium sphere. (a) Rotationally invariant content of the sample. Diagonal t bands of (b) $t \leq 14$, (c) $15 \leq t \leq 35$, (d) $36 \leq t \leq 50$, (e) $51 \leq t \leq 78$, (f) $79 \leq t \leq 110$, (g) $111 \leq t \leq 119$, and (h) $120 \leq t$. (i) Superposition of wavefronts in (a) to (h) and its 2D MTF is shown on the right.

Figure 5.19 shows the error in the calculation of MTF based on the product of MTFs of wavefronts in Figure 5.18. The error in the calculation MTF based on Eqn. 5.8 is 0.4 % that is less than one percent of the volume under the 2D MTF surface of the sum wavefront.

$$\left[MTF_1 \times MTF_2 \times MTF_3 \times MTF_4 \times MTF_5 \times MTF_6 \times MTF_7 \times MTF_8\right] - \left[MTF_{1+2+3+4+5+6+7+8} \times MTF_{diff}^{7}\right]$$

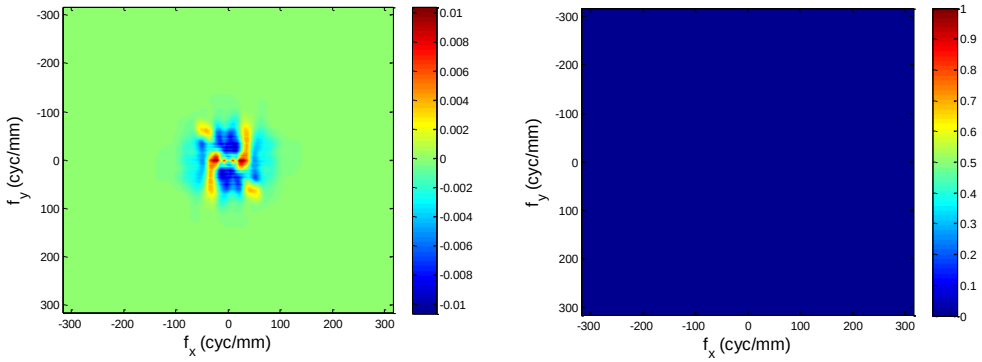


Figure 5.19 (a) Error in the approximation of MTF obtained by subtracting the MTF of the superposition wavefronts in Figure 5.18 and the product of MTFs of all bands. (b) The MTF error when we change the color scale to [0 1]. There is no visible pattern in the re-scaled MTF error and the difference is negligible.

Figure 5.20 shows the azimuthal bands of the germanium sample. Similar to Figure 5.18, the primary astigmatism term has been removed due to its large RMS before carrying out the filtering. The wavefront in Figure 5.20 (f) is the sum of the filtered bands in Figure 5.20 a-e. The MTF of the sum wavefront is also equal to the product of the MTFs of its components that are shown in the second row of Figure 5.20. Figure 5.21 shows the error in the calculation of MTF by calculating the difference between the MTF of the sum wavefront and the product of MTFs of all the bands. We have accounted for the diffraction-limited MTF in simulating the MTF error in Figure 5.17. The volume MTF error based on Eqn.5.8 and is equal to ~0.9%, which is less than one percent of the area MTF of the sum wavefront.

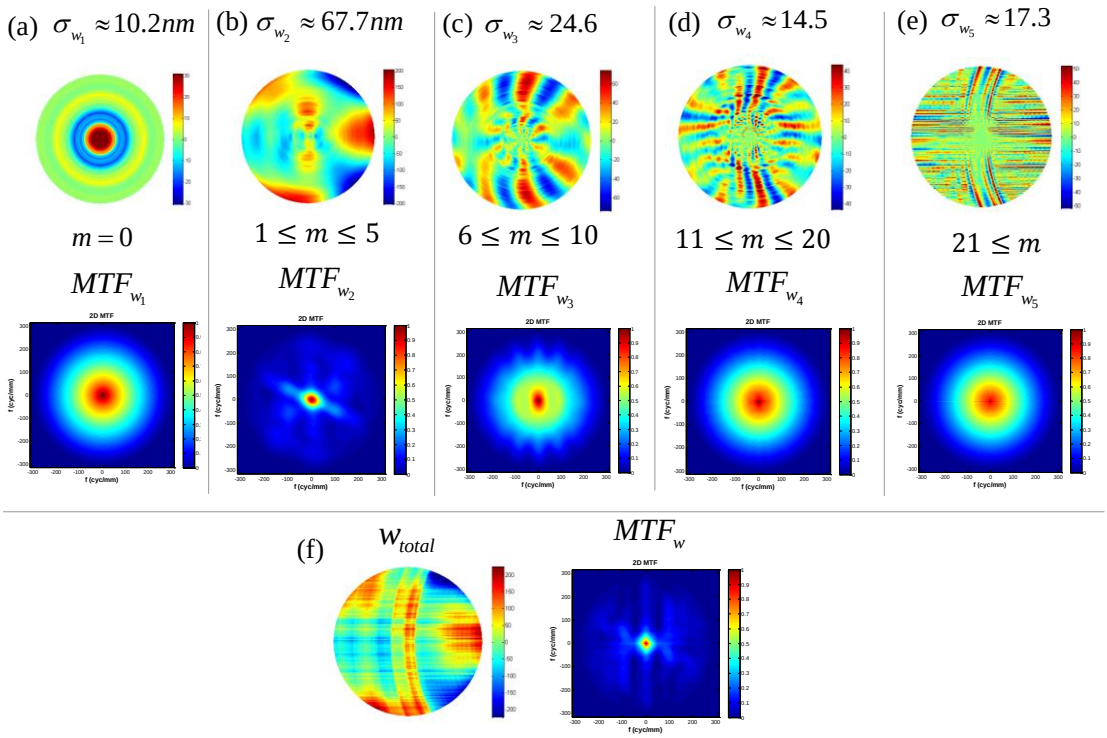


Figure 5.20 (a) Rotationally invariant bands of the geranium sample with RMS of 10.2 nm. (b) Selected azimuthal band of $1 \le m \le 5$ with the RMS of 67.7 nm. (c) Azimuthal band of $6 \le m \le 10$ with the RMS of 24.6 nm. (d) Azimuthal band of $11 \le m \le 20$ with the RMS of 14.5 nm. (e) Azimuthal band of $21 \le m$ with the RMS of 17.3 nm. (f) Superposition of bands in (a) to (e) and its 2D MTF.

$$\left[MTF_1 \times MTF_2 \times MTF_3 \times MTF_4 \times MTF_5\right] - \left[MTF_{1+2+3+4+5} \times MTF_{diff}^4\right]$$

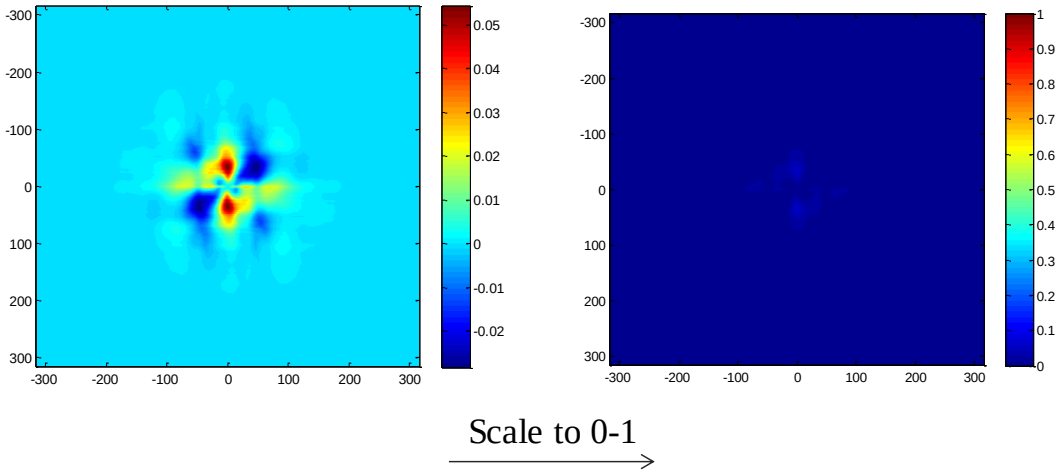


Figure 5.21 (a) Error in the approximation of MTF for the filtered bands of Figure 5.20. (b) The MTF error when we change the color scale to [0 1]. The difference is negligible.

Here we show an example of the application of the linear systems theory of MTF on isolated bands of a wavefront based on sources of fabrication errors. In Figure 5.22, different filtered bands of the IRG sphere are shown. Figure 5.22 (a) shows all the ring-like structures, (b), (c) and (d) are filtered bands of $0 \le t \le 10$, $11 \le t \le 50$, and $51 \le t$ that also exclude all of the rotationally invariant terms. This part was fabricated with diamond raster milling and the direction of the feed is obvious in Figures 5.22 (c) and (d). Figure 5.22 (a) shows the rotationally invariant content of the surface. These errors are caused by misalignment of the diamond cutting tool (incorrect center and not exactly perpendicular to the shank [124]). Figure 5.22 (b) shows large-scale raster patterns (maximum five cycles across aperture). These errors are the consequence of three different sources that are mainly thermal. These include i) thermal oscillations in the milling spindle chiller, ii) thermal oscillations in the work-holding spindle chiller iii) uneven thermal distortion due to the handling of the part during mounting on the machine [124]. The raster-like texture in Figures 5.22 (c) and (d) are mainly caused by the step-over in the programmed tool path. In Figure 5.22 (d), one can see locations of two dust particles near the edge of the part. Some of the high-frequency errors in Figure 5.22 (d) are caused by the tool vibrations due to sharp changes in direction and velocity of the toolpath [124].

The second row in Figure 5.22 shows the 2D MTF of the filtered bands. Depending on the application, one could specify the maximum geometrical MTF drop in a specific frequency range if resolving a specific spatial frequency is needed such as in lithography tools. For example, if the specification requires that the maximum geometrical MTF drop for the superposition wavefront at the spatial frequency of $f_x = 12.16\ 1/mm$ and $f_y = 8.1\ 1/mm$ to be no smaller than 0.65 and equal budgets are required for rotationally-

invariant terms and raster-like patterns, then to meet the specification, both geometrical MTFs should be no smaller than $\sqrt[2]{MTF_{Spec}^{Geom}|_{f_x,f_y}}$ at those spatial frequencies. If we have N number of bands, for equal budget specification for each band, the geometrical MTF of each band should be no smaller than $\sqrt[N]{MTF_{Spec}^{Geom}|_{f_x,f_y}}$, where N is the number of bands. In this example, to meet the specification, both the rotationally invariant band and the non-rotationally invariant band should have geometrical MTFs of no smaller than 0.81 at the specified spatial frequencies. Geometrical MTF of rotationally invariant terms shown in Figure 5.23 (a) is 0.73 at $f_x = 12.16\ 1/mm$ and $f_y = 8.1\ 1/mm$ that does not meet the specification. The second row in Figure 5.23 shows the cross section of the 2D geometrical MTF at $f_y = 8.1\ 1/mm$. To meet our specification, fixing the centering of the cutting tool should be a priority. The diamond cutting tool should be centered properly and it should be perpendicular to the shank. Non-rotationally invariant terms as an aggregate meet the specification. However, if we had to specify each band separately allowing equal budget, the geometrical MTF of each band should be no smaller than 0.94 at those frequencies. The filtered band in (b) with $MTF_{W_2}^{Geom} = 0.99$ meets the specification. The filtered band in (c), W_3, does not meet the specification since $MTF_{W_3}^{Geom} = 0.93$ but it is very close to the specified value. The filtered band in (d) with $MTF_{W_4}^{Geom} = 0.94$ meets the specification. The $MTF_{W_4}^{f_x,f_y} = 0.889$ and depending on the specified uncertainty in the reporting of MTF, may or may not meet the specification.

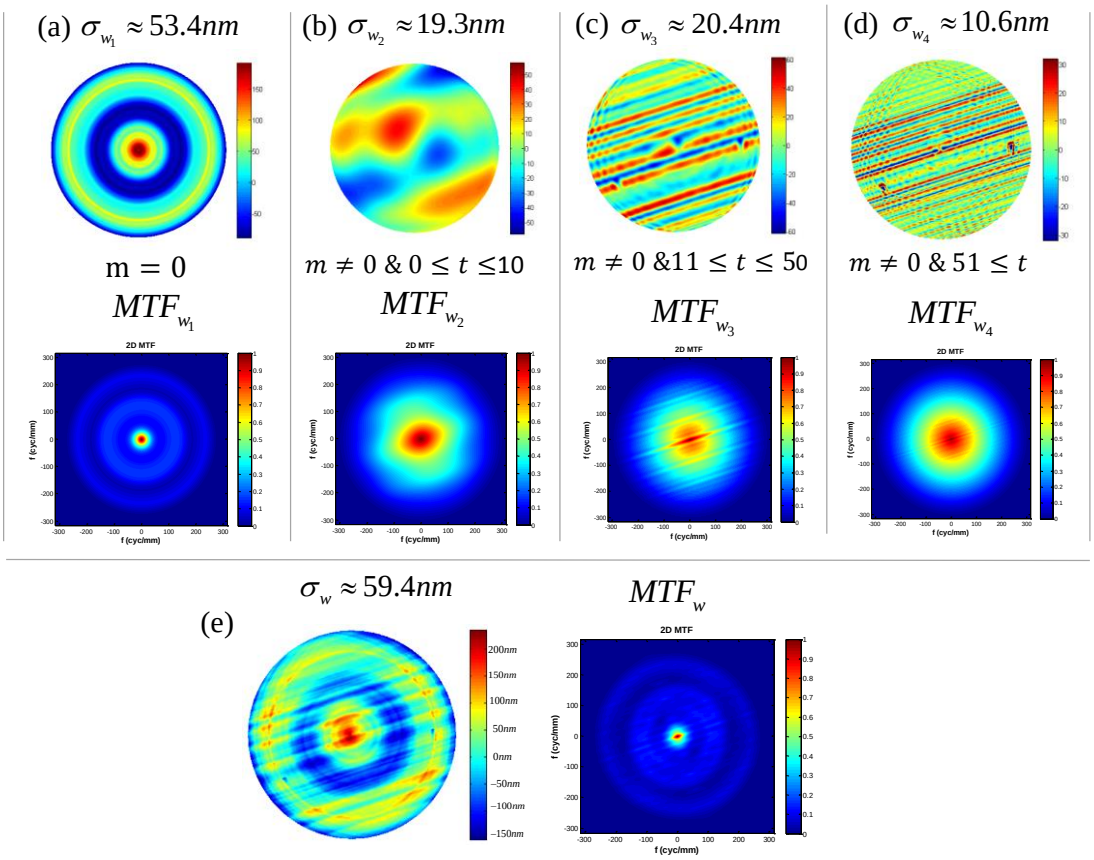


Figure 5.22 Orthogonal bands of the IRG 26 sphere. (a) Rotationally-invariant content of the IRG 26 sphere. (b) Diagonal band of $0 \leq t \leq 10$ with the rotationally-invariant content removed. (c) Diagonal band of $11 \leq t \leq 50$ excluding the rotationally invariant terms. (d) Diagonal band of $51 \leq t$ with the rotationally-invariant terms removed. (e) Superposition of wavefronts in (a) to (d) and it's 2D MTF.

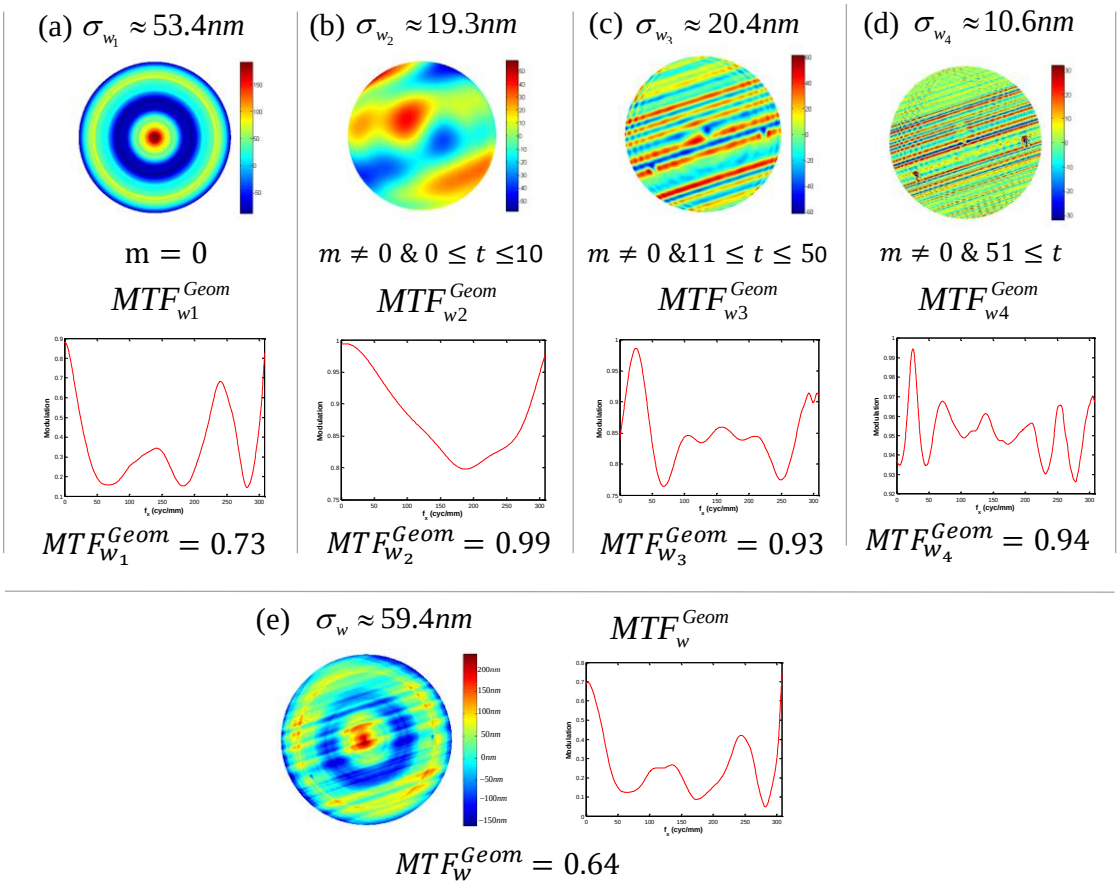


Figure 5.23 Orthogonal bands of the IRG 26 shown in the first row and their cross-sectional MTF at $f_y = 8.1\ 1/mm$ in the second row. (a) Rotationally-invariant content of the IRG 26 sphere. (b) Diagonal bands of $0 \leq t \leq 10$ with the rotationally-invariant content removed. (c) Diagonal band of $11 \leq t \leq 50$ excluding the rotationally invariant terms. (d) Diagonal band of $51 \leq t$ with the rotationally-invariant terms removed. (e) Superposition of wavefronts in (a) to (d) and its cross-sectional MTF at $f_y = 8.1\ 1/mm$.

CHAPTER 6 : CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

In Chapter one, we reviewed the importance of characterizing the mid-spatial frequency texture due to its effect on the final image. We also showed different spatial frequency regions of form (Figure), mid-spatial frequency (waviness), roughness, and their definitions from the literature along with an example of a measured surface with different frequency regions separated. In 1.3, we introduced different mathematical methods such as power spectral density (PSD) and Zernike polynomial analysis for characterization of mid-spatial frequencies.

In Chapter two, we compared these mathematical methods against chosen evaluation criteria. Our evaluation criteria were the amount of pre-processing, the uniqueness of the result (related to orthogonality), connection to fabrication choices, dependence on aperture geometry, ability to generalize to full frequency range, connection to optical performance and level of intuitiveness. Zernike polynomials are orthogonal over a unit circle and for characterizing circular aberrations with these polynomials, there is no need for pre-processing the data.

In Chapter three, we showed the generation of very large orders of Zernike polynomials using recursive formulas to avoid round-off errors. The Zernike polynomials generated extend to more than 45,000 terms, and they represent a unique possibility for characterizing MSF errors on freeform surfaces. The results of fitting pure sinusoidal patterns to high

orders of Zernike polynomials reveal the signatures of MSF texture in the fit coefficient maps. We showed that for raster-like patterns the values of n_f and m where the coefficients are largest happen at unique values corresponding to the length-scale of the periodicity. For rotationally invariant patterns, the peak of the coefficient distribution occurs at rotationally invariant orders where the azimuthal order m is zero. For spokes patterns, the peak of the distribution is at an azimuthal order m, where m equals the number of azimuthal periods in the map. We then evaluated the effectiveness of Zernike characterization in identifying the length scale, amplitude and, type of MSF by comparing this analysis to PSD. Having built intuition on how to interpret Zernike fit coefficients maps, we showed fit coefficient maps of real mid-spatial frequency texture where we can observe and interpret signatures of different fabrication processes. We showed that the RMS of fit residuals decreases as the number of basis elements goes up, and this supports the conclusion that the fit coefficients are numerically stable even for these very large orders. However, the rate of drop in RMS residual decreases, in part due to choice in the ordering of the fit. At very extreme orders, the true polynomial map has more than one zero-crossing oscillation within each pixel around the edge of the aperture. Real measured surface data are discrete and have limited array size. Therefore, we cannot increase the number of polynomial basis arbitrarily.

In Chapter four, we discuss optical performance metrics of imaging systems in the context of MSF surface error. The primary metrics used are the point spread function (PSF), the modulation transfer function (MTF) and the Strehl ratio. We considered these metrics for both synthetic and real MSF surfaces using both commercial software (Mx) and MATLAB. We showed that raster-like patterns on an optical surface of a sinusoidal surface

and of a cusp-shaped surface with the same PV surface error have different MTF responses. Pure sinusoidal MSF textures on an optical surface degrade the optical performance more than the cusp-shaped texture of the same PV. We also showed that MTF loss has a non-linear relationship with PV, and we recommend using an optical path difference that embodies the impact of both refractive index and incident wavelength for optical performance analysis and specification purposes.

In Chapter five, we showed how Zernike fit coefficients of a surface can be used to predict the Strehl ratio. We showed that the Strehl ratio approximation $\exp(-k^2\sigma^2)$ is valid for a large portion of the Strehl ratio curve, as well as large order Zernike polynomial terms. The nominal orthogonality property of Zernike polynomials provides an effective tool in estimating the Strehl ratio of the whole surface based on the Strehl ratios of each filtered band. The limits of validity of the Strehl ratio approximation for MSF texture and large polynomial orders was analyzed using some examples. We also showed that one can take advantage of the linear systems properties of MTF and find the MTF of the entire surface based on the MTF of individual filtered bands. The nominal orthogonality properties of Zernike polynomials, along with linear systems formulation of MTF, provide us with a powerful tool to connect surface characterization metrics, such as filtering based on Zernike polynomials, with design and performance metrics, such as MTF. Knowing the optical performance degradation contribution due to individual bands and their relationship to the degradation of the whole surface enables designers and metrologists to specify MSF due to different fabrication sources.

6.2 Assumptions and validity of our optical performance analysis

In order to describe the interaction of light with surface texture completely, a rigorous

coupled wave analysis (RCWA) is needed. In the optical performance analysis in this dissertation, we have used scalar diffraction theory. The treatment of mid-spatial frequency texture as grating structures and using scalar diffraction theory sets a limit on the surface heights and feature sizes of the surface for which the approximation is valid. In order for scalar diffraction theory to be valid, the spatial periods on the sample must be larger than Λ_{min} where $\Lambda_{min} = \frac{\lambda}{n}$. Harvey *et al.* use an extended diffraction theory for small feature sizes (compared to the wavelength) and large scattering angles [119, 123].

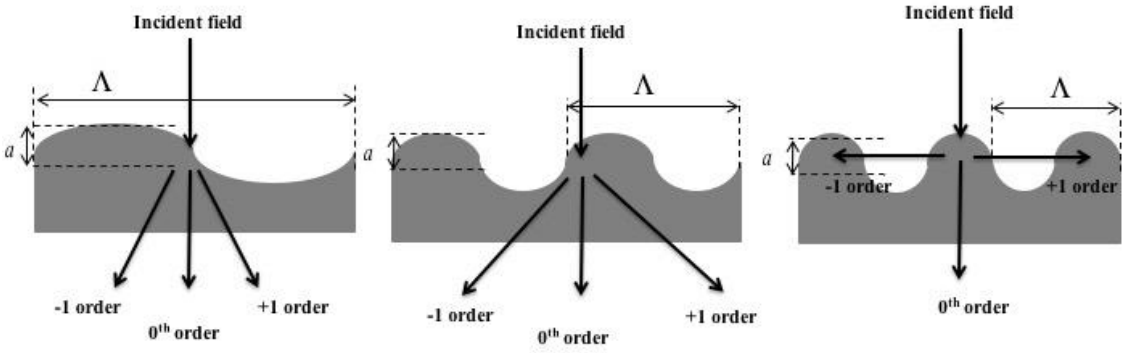


Figure 6.1 Mid-spatial frequency errors are grating-like structures. As the surface wavelength gets smaller, the diffraction angle increases. When $\Lambda<\Lambda_{min}$ the diffracted orders become evanescent waves.

Another assumption for this analysis is that the light interacts with the surface once. As a result, we ignore the propagation of waves inside the grating structure and ignore implications for the effective medium theory. We use Rayleigh-Sommerfeld diffraction theory for our analysis, which is based on scalar diffraction theory. The Rayleigh-Sommerfeld diffraction theory is valid when the propagation medium is non-magnetic, linear, isotropic and homogeneous, which allows for the expression of Maxwell's equations as scalar wave equations (scalar diffraction theory). Then the Sommerfeld radiation condition applies here, which guarantees that we are only dealing with outgoing waves and

the integral over S_2 in Figure 6.2 will be zero.

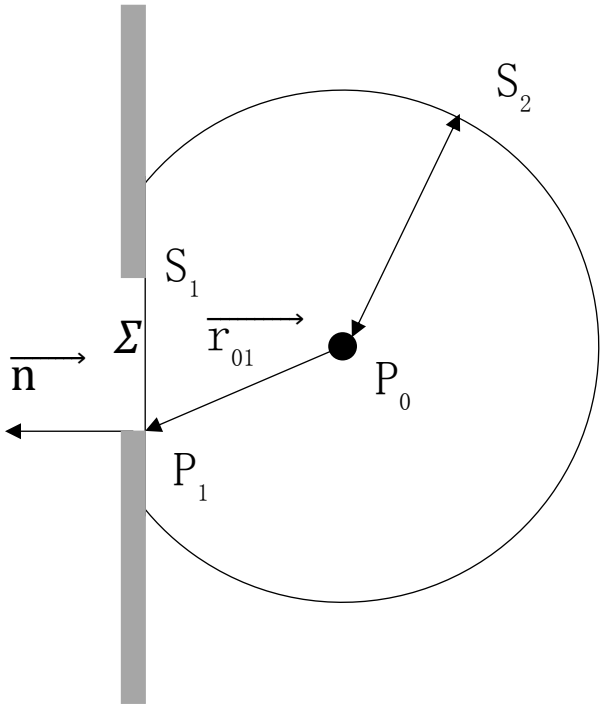


Figure 6.2 Kirchholf formulation of diffraction [25].

Using Rayleigh-Sommerfeld theory, the field at point p_0 can be written as

$$U(p_0) = \frac{A}{j\lambda}\iint \frac{\exp\left[jk(r_{21}+r_{01})\right]}{r_{21}r_{01}}\cos(\vec{n},\vec{r}_{01})ds \tag{6.1}$$

where $\cos(\vec{n},\vec{r}_{01})$ is the obliquity factor, and for the case of an infinitely distant point source with plane wave illumination (which is the case for calculating point spread function and modulation transfer function), the obliquity factor is equal to 1.

Another assumption is that $r_{01} \gg \lambda$, i.e. the distance between the observation point and the diffracting surface is much larger than the incident wavelength.

6.3 Future work

As described in the previous chapters, measurements of a surface are limited by the instrument transfer function and the Nyquist sampling frequency. As a result, we cannot fit

the data to arbitrary orders of Zernike polynomials. Residual maps of MSF surfaces reveal clear structured texture even after removal of very large orders. Figure 6.3 shows the fit residual maps of measured surfaces after removal of 200 orders of Zernike polynomials. For small wavelengths and high-performance applications, this residual texture can be significant and further degrade the optical performance. Thus, a Zernike representation of MSF errors is unable to adequately capture the impact of the texture and provide a reliable prediction of optical performance. When we go to higher orders of Zernike polynomials, the zero crossings of the polynomials at the aperture edge and near the center become more closely spaced than the pixels in the data. Interpolating our data to a non-uniform grid such as an edge-clustered- recommended in [107]-, or a non-uniform grid depicted in [49] would enable robust fitting to even higher orders of polynomials.

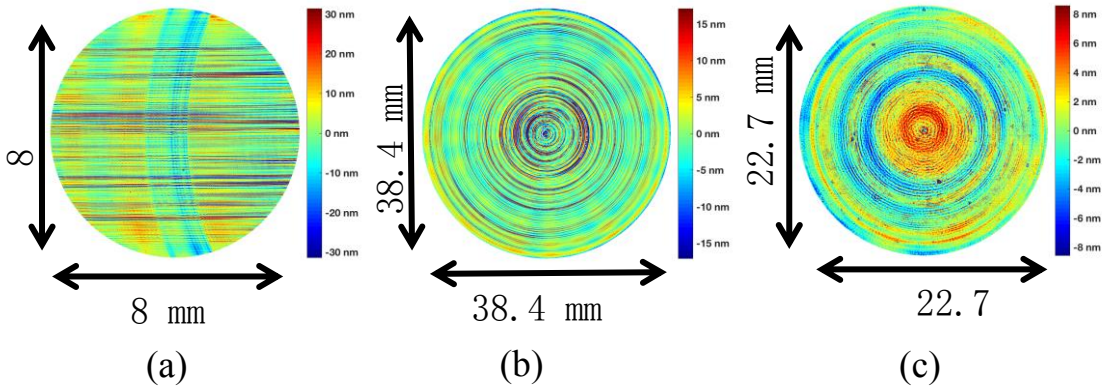


Figure 6.3 The residual maps of three surfaces after removal of 200 orders of Zernike polynomials. (a) The residual map of germanium sphere with RMS error of 10.4 nm. (b) Residual map of diamond turned aluminum part with RMS error of 5.7 nm. (c) Residual map of SiC sphere with RMS error of 2.8 nm.

Measurements from multiple instruments with varying resolution can be combined to provide the whole surface map with a greater data density. L. He *et al.* have combined area structure function calculations from different instruments (a Fizeau phase shifting interferometer and a Zygo ZeGage environmentally tolerant coherence scanning

interferometer (CSI)) [59]. Sub-aperture stitching ideas were first developed to test large optical flats and spheres [125]. Thunen and Kwon have developed algorithms based on Zernike polynomials to construct full aperture data error from the measurement of single sub-apertures [94, 126]. The main problem in the accuracy of sub-aperture stitching methods are the uncertainties in the relative piston and tilt of sub-apertures. These uncertainties can be estimated and then added to the sub-aperture data before performing the analysis. Other methods exist that require overlapping regions among sub-apertures, and the piston and tilt are estimated by least square fitting to the difference of overlapping data [127-129]. Murphy *et al.* extended sub-aperture stitching to the measurement of mild aspheres [130]. Tricard *et al.* have incorporated variable optical null (VON) in sub-aperture stitching interferometry for measurement of high departure aspheres with up to 1000 waves of departure from the best-fit sphere. VON technology uses configurable auxiliary optics to generate aberrations that match with all different sub-apertures [131]. Predicting the optical performance of full aperture data from sub-aperture data based on Zernike polynomial analysis can also be valuable for future research in freeform optics.

REFERENCES

[1] K.P. Thompson and J.P. Rolland,"Freeform optical surfaces: a revolution in imaging optical design,"*Optics and Photonics News* 23 (6), 30-35 (2012).

[2] D. Cheng, et al.,"Design of an optical see-through head-mounted display with a low f-number and large field of view using a freeform prism,"*Applied optics* 48 (14), 2655-2668 (2009).

[3] Y. Ding, et al.,"Freeform LED lens for uniform illumination,"*Optics Express* 16 (17), 12958-12966 (2008).

[4] H. Ries and J. Muschaweck,"Tailored freeform optical surfaces,"*JOSA A* 19 (3), 590-595 (2002).

[5] K. Garrard, et al., "Design tools for freeform optics." Optics & Photonics 2005. International Society for Optics and Photonics p. 58740A-58740A-11 (2005).

[6] F. Fang, et al.,"Manufacturing and measurement of freeform optics,"*CIRP Annals-Manufacturing Technology* 62 (2), 823-846 (2013).

[7] K.P. Thompson, P. Benítez, and J.P. Rolland,"Freeform optical surfaces: Report from OSA's first incubator meeting,"*Optics and Photonics News* 23 (9), 32-37 (2012).

[8] F. Tinker and K. Xin, "Fabrication of SiC aspheric mirrors with low mid-spatial error." SPIE Optical Engineering+ Applications. International Society for Optics and Photonics p. 88370M-88370M-7 (2013).

[9] C.R. Dunn and D.D. Walker,"Pseudo-random tool paths for CNC sub-aperture polishing and other applications,"*Optics express* 16 (23), 18942-18949 (2008).

[10] D.M. Aikens, C.R. Wolfe, and J.K. Lawson, "Use of power spectral density (PSD) functions in specifying optics for the National Ignition Facility." International Conferences on Optical Fabrication and Testing and Applications of Optical Holography. International Society for Optics and Photonics p. 281-292 (1995).

[11] J.K. Lawson, et al., "Specification of optical components using the power spectral density function." SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics p. 38-50 (1995).

[12] R.W. Boyd, S.G. Lukishova, and Y.R. Shen, *Self-focusing: Past and Present:*

Fundamentals and Prospects. Vol. 114. Springer Science & Business Media (2008).

[13] B. Van Wonterghem, et al.,"System description and initial performance results for Beamlet,"*ICF quarterly Report* 5 (1), 1-17 (1994).

[14] J.L. Bentley and C. Olson, "Field guide to lens design." SPIE (2012).

[15] J.M. Tamkin, T.D. Milster, and W. Dallas,"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).

[16] R.N. Youngworth and B.D. Stone,"Simple estimates for the effects of mid-spatial-frequency surface errors on image quality,"*Applied optics* 39 (13), 2198-2209 (2000).

[17] L.L. Deck and C. Evans, "High performance Fizeau and scanning white-light interferometers for mid-spatial frequency optical testing of free-form optics." Optics & Photonics 2005. International Society for Optics and Photonics p. 59210A-59210A-8 (2005).

[18] M. Chandhok, et al., "Comparison of techniques to measure the point spread function due to scatter and flare in EUV lithography systems." Microlithography 2004. International Society for Optics and Photonics p. 854-860 (2004).

[19] T. Wang, et al.,"Controlling mid-spatial frequency errors in magnetorheological jet polishing with a simple vertical model,"*Applied optics* 54 (21), 6433-6440 (2015).

[20] P. De Groot and X.C. De Lega, Interpreting interferometric height measurements using the instrument transfer function, in *Fringe 2005*. Springer. p. 30-37 (2006).

[21] D.M. Sykora and P. De Groot, "Instantaneous measurement Fizeau interferometer with high spatial resolution." SPIE Optical Engineering+ Applications. International Society for Optics and Photonics p. 812610-812610-10 (2011).

[22] L. Deck,"High precision interferometer for measuring mid-spatial frequency departure in free form optics,"*SPIE Technical Digest, TD04 pp. TD040M* (2007).

[23] D. Aikens, J.E. Degroote, and R.N. Youngworth, "Specification and control of mid-spatial frequency wavefront errors in optical systems." Optical Fabrication and Testing. Optical Society of America p. OTuA1 (2008).

[24] J.M. Tamkin and T.D. Milster,"Effects of structured mid-spatial frequency surface errors on image performance,"*Applied optics* 49 (33), 6522-6536 (2010).

[25] J.W. Goodman, *Introduction to Fourier optics*. Roberts and Company Publishers (2005).

[26] R. Williamson, "Field Guide to Optical Fabrication." Spie (2011).

[27] C.J. Evans,"PVr—a robust amplitude parameter for optical surface specification,"*Optical Engineering* 48 (4), 043605-043605-8 (2009).

[28] J.E. Greivenkamp, *Field guide to geometrical optics*. Vol. 1. SPIE Press Bellingham, Washington (2004).

[29] J.C. Wyant and K. Creath,"Basic wavefront aberration theory for optical metrology,"*Applied optics and optical engineering* 11 (s 29), 2 (1992).

[30] H. Davies,"The reflection of electromagnetic waves from a rough surface,"*Proceedings of the IEE-Part IV: Institution Monographs* 101 (7), 209-214 (1954).

[31] N. Bruce and J. Dainty,"Multiple scattering from random rough surfaces using the Kirchhoff approximation,"*Journal of modern optics* 38 (3), 579-590 (1991).

[32] E.L. Church and P.Z. Takacs, "Light scattering from non-Gaussian surfaces." SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics p. 91-107 (1995).

[33] A. Kotha and J.E. Harvey, "Scattering effects of machined optical surfaces." SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics p. 54-65 (1995).

[34] J. Krill and R. Farrell,"Comparisons between variational, perturbational, and exact solutions for scattering from a random rough-surface model,"*Journal of the Optical Society of America (1917-1983)* 68, 768-774 (1978).

[35] S. Schröder and A. Duparré, "Finish assessment of complex surfaces by advanced light scattering techniques." Optical Systems Design. International Society for Optics and Photonics p. 71020F-71020F-9 (2008).

[36] J.C. Stover,"Roughness characterization of smooth machined surfaces by light scattering,"*Applied Optics* 14 (8), 1796-1802 (1975).

[37] D.Y. Wang, R.E. English Jr, and D.M. Aikens, "Implementation of ISO 10110 optics drawing standards for the National Ignition Facility." SPIE's International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics p. 502-508 (1999).

[38] J.K. Lawson, et al., "Power spectral density specifications for high-power laser systems." Optical Instrumentation & Systems Design. International Society for Optics and Photonics p. 345-356 (1996).

[39] J.H. Campbell, et al., "NIF optical materials and fabrication technologies: an overview." Lasers and Applications in Science and Engineering. International Society for Optics and Photonics p. 84-101 (2004).

[40] J.M. Elson and J.M. Bennett,"Calculation of the power spectral density from surface profile data,"*Applied Optics* 34 (1), 201-208 (1995).

[41] R.N. Youngworth, B.B. Gallagher, and B.L. Stamper, "An overview of power spectral density (PSD) calculations." Optics & Photonics 2005. International Society for Optics and Photonics p. 58690U-58690U-11 (2005).

[42] P.Z. Takacs, "Minimizing artifacts in analysis of surface statistics." Optical Fabrication and Testing. Optical Society of America p. OTu1D. 5 (2012).

[43] P.Z. Takacs, "Standardization of methods for extracting statistics from surface profile measurements." SPIE NanoScience+ Engineering. International Society for Optics and Photonics p. 917309-917309-10 (2014).

[44] E. Church and P. Takacs,"BASIC program for power spectrum estimation," (1994).

[45] V. Lakshminarayanan and A. Fleck,"Zernike polynomials: a guide,"*Journal of Modern Optics* 58 (7), 545-561 (2011).

[46] E.P. Goodwin and J.C. Wyant, "Field Guide to Interferometric Optical Testing." SPIE (2006).

[47] J.S. Loomis,"FRINGE User's manual,"*D. Anderson, Fringe Manual Version* 3 (1976).

[48] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. CUP Archive (2000).

[49] G. Forbes,"Fitting freeform shapes with orthogonal bases,"*Optics express* 21 (16), 19061-19081 (2013).

[50] G. Forbes,"Shape specification for axially symmetric optical surfaces,"*Optics express* 15 (8), 5218-5226 (2007).

[51] C. Menke and G. Forbes,"Optical design with orthogonal representations of rotationally symmetric and freeform aspheres,"*Advanced Optical Technologies* 2 (1), 97-109 (2013).

[52] A.N. Kolmogorov, "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers." Dokl. Akad. Nauk SSSR. JSTOR p. 301-

305 (1941).

[53] D.L. Fried,"Statistics of a geometric representation of wavefront distortion,"*JOSA* 55 (11), 1427-1435 (1965).

[54] J.H. Burge, et al., "Design and analysis for interferometric measurements of the GMT primary mirror segments." SPIE Astronomical Telescopes+ Instrumentation. International Society for Optics and Photonics p. 62730M-62730M-12 (2006).

[55] A.M. Hvisc and J.H. Burge, "Structure function analysis of mirror fabrication and support errors." Optical Engineering+ Applications. International Society for Optics and Photonics p. 66710A-66710A-10 (2007).

[56] T. Thomas, B.-G. Rosen, and N. Amini,"Fractal characterisation of the anisotropy of rough surfaces,"*Wear* 232 (1), 41-50 (1999).

[57] R. Sayles and T. Thomas,"The spatial representation of surface roughness by means of the structure function: a practical alternative to correlation,"*Wear* 42 (2), 263-276 (1977).

[58] L. He, C.J. Evans, and A. Davies,"Spatial content analysis for precision surfaces with the area structure function,"*Surface Topography: Metrology and Properties* 2 (1), 014010 (2013).

[59] L. He, C.J. Evans, and A. Davies,"Optical surface characterization with the area structure function,"*CIRP Annals-Manufacturing Technology* 62 (1), 539-542 (2013).

[60] L. He, C.J. Evans, and A. Davies,"Two-quadrant area structure function analysis for optical surface characterization,"*Optics express* 20 (21), 23275-23280 (2012).

[61] D.J. Whitehouse, *Handbook of surface and nanometrology*. CRC press (2010).

[62] B. Muralikrishnan and J. Raja, *Computational surface and roundness metrology*. Springer Science & Business Media (2008).

[63] D.J. Whitehouse and J. Archard, "The properties of random surfaces of significance in their contact." Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society p. 97-121 (1970).

[64] L. He, Optical surface characterization with the structure function. The University of North Carolina at Charlotte (2013).

[65] B. Josso, D.R. Burton, and M.J. Lalor,"Frequency normalised wavelet transform for surface roughness analysis and characterisation,"*Wear* 252 (5), 491-500 (2002).

[66] I. Daubechies, *Ten lectures on wavelets*. Vol. 61. SIAM (1992).

[67] J.S. Walker, *A primer on wavelets and their scientific applications*. CRC press (2008).

[68] N. Kingsbury,"Complex wavelets for shift invariant analysis and filtering of signals,"*Applied and computational harmonic analysis* 10 (3), 234-253 (2001).

[69] S.G. Mallat,"A theory for multiresolution signal decomposition: the wavelet representation,"*Pattern Analysis and Machine Intelligence, IEEE Transactions on* 11 (7), 674-693 (1989).

[70] P. Jester, C. Menke, and K. Urban,"Wavelet methods for the representation, analysis and simulation of optical surfaces,"*IMA Journal of Applied Mathematics* 77 (4), 495-515 (2012).

[71] X. Jiang, P. Cooper, and P.J. Scott, "Freeform surface filtering using the diffusion equation." Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society p. 841-859 (2011).

[72] H.S. Abdul-Rahman, X.J. Jiang, and P.J. Scott,"Freeform surface filtering using the lifting wavelet transform,"*Precision Engineering* 37 (1), 187-202 (2013).

[73] W. Sweldens,"The lifting scheme: A custom-design construction of biorthogonal wavelets,"*Applied and computational harmonic analysis* 3 (2), 186-200 (1996).

[74] Iso, 10110 Optics and optical instruments – Preparation of drawings for optical elements and systems, in Part 8: Surface texture. (1997).

[75] G. Erdei, G. Szarvas, and E. Lorincz, "Tolerancing surface accuracy of aspheric lenses used for imaging purposes." Optical Systems Design. International Society for Optics and Photonics p. 718-728 (2004).

[76] J.R. Rogers, "Slope error tolerances for optical surfaces." Presentation at OptiFab Conference,, SPIE TD04-4. (2007).

[77] K. Prabhu, *Window functions and their applications in signal processing*. CRC Press (2013).

[78] F.J. Harris,"On the use of windows for harmonic analysis with the discrete Fourier transform,"*Proceedings of the IEEE* 66 (1), 51-83 (1978).

[79] S.D. Stearns and D.R. Hush, *Digital Signal Processing with Examples in MATLAB®*. CRC press (2002).

[80] MetroProX. Zygo Corp (2015).

[81] W. Swantner and W.W. Chow,"Gram–Schmidt orthonormalization of Zernike polynomials for general aperture shapes,"*Applied optics* 33 (10), 1832-1837 (1994).

[82] J. Ye, et al.,"Comparative assessment of orthogonal polynomials for wavefront reconstruction over the square aperture,"*JOSA A* 31 (10), 2304-2311 (2014).

[83] F. Liu, et al.,"Analyzing optics test data on rectangular apertures using 2-D Chebyshev polynomials,"*Optical Engineering* 50 (4), 043609-043609-8 (2011).

[84] G.W. Forbes, "Detrending on Steroids... MSF Characterization via Freeform Polynomials." Freeform Optics. Optical Society of America p. FM2B. 1 (2013).

[85] R. Bracewell,"The Fourier Transform and Its Applications," (2000).

[86] E.C. Kintner,"On the mathematical properties of the Zernike polynomials,"*Journal of Modern Optics* 23 (8), 679-680 (1976).

[87] G. Forbes,"Robust and fast computation for the polynomials of optics,"*Optics express* 18 (13), 13851-13862 (2010).

[88] C.J. Evans, et al.,"Visualization of surface figure by the use of Zernike polynomials,"*Applied optics* 34 (34), 7815-7819 (1995).

[89] J.W. Goodman, *Statistical optics*. John Wiley & Sons (2015).

[90] V.N. Mahajan,"Zernike annular polynomials for imaging systems with annular pupils,"*JOSA* 71 (1), 75-85 (1981).

[91] R. Upton and B. Ellerbroek,"Gram–Schmidt orthogonalization of the Zernike polynomials on apertures of arbitrary shape,"*Optics letters* 29 (24), 2840-2842 (2004).

[92] E. Sidick, "Power spectral density specification and analysis of large optical surfaces." SPIE Europe Optical Metrology. International Society for Optics and Photonics p. 73900L-73900L-12 (2009).

[93] I. Kaya, K.P. Thompson, and J.P. Rolland,"Comparative assessment of freeform polynomials as optical surface descriptions,"*Optics express* 20 (20), 22683-22691 (2012).

[94] J.G. Thunen and O.Y. Kwon, "Full aperture testing with subaperture test optics." 26th Annual Technical Symposium. International Society for Optics and Photonics p. 19-27 (1983).

[95] W.W. Chow and G.N. Lawrence, "Analysis of subaperture testing data." 27th

Annual Techincal Symposium. International Society for Optics and Photonics p. 99-105 (1984).

[96] J.E. Harvey and A.K. Thompson, "Scattering effects from residual optical fabrication errors." International Conferences on Optical Fabrication and Testing and Applications of Optical Holography. International Society for Optics and Photonics p. 155-174 (1995).

[97] V. Mahajan, "Optical Imaging and Aberrations, Part II. Wave Diffraction Optics." SPIE (2011).

[98] J.M. Bennett,"Measurement of the rms roughness, autocovariance function and other statistical properties of optical surfaces using a FECO scanning interferometer,"*Applied Optics* 15 (11), 2705-2721 (1976).

[99] E.L. Church and P.Z. Takacs,"Specification of surface figure and finish in terms of system performance,"*Applied Optics* 32 (19), 3344-3353 (1993).

[100] G. Forbes, "Never-ending struggles with mid-spatial frequencies." SPIE Optical Metrology. International Society for Optics and Photonics p. 95251B-95251B-10 (2015).

[101] A. Bhatia and E. Wolf, "On the circle polynomials of Zernike and related orthogonal sets." Mathematical Proceedings of the Cambridge Philosophical Society. Cambridge Univ Press p. 40-48 (1954).

[102] V.N. Mahajan,"Zernike circle polynomials and optical aberrations of systems with circular pupils,"*Applied optics* 33 (34), 8121-8124 (1994).

[103] G. Szeg, *Orthogonal polynomials*. Vol. 23. American Mathematical Soc. (1939).

[104] R.J. Noll,"Zernike polynomials and atmospheric turbulence*,"*JOsA* 66 (3), 207-211 (1976).

[105] F.J. Smith,"An algorithm for summing orthogonal polynomial series and their derivatives with applications to curve-fitting and interpolation,"*Mathematics of Computation* 19 (89), 33-36 (1965).

[106] M. Abramowitz and I.A. Stegun,"Handbook of mathematical functions,"*Applied mathematics series* 55, 62 (1966).

[107] I. Kaya, K.P. Thompson, and J.P. Rolland,"Edge clustered fitting grids for φ-polynomial characterization of freeform optical surfaces,"*Optics express* 19 (27), 26962-26974 (2011).

[108] Available from: https://pypi.python.org/pypi/Scikit-Qfit.

[109] J. Wang and D.E. Silva,"Wave-front interpretation with Zernike polynomials,"*Applied optics* 19 (9), 1510-1518 (1980).

[110] R.J. Mathar,"Third order Newton's method for Zernike polynomial zeros,"*arXiv preprint arXiv:0705.1329* (2007).

[111] V.N. Mahajan,"Strehl ratio for primary aberrations in terms of their aberration variance,"*JOSA* 73 (6), 860-861 (1983).

[112] J. Marioge and S. Slansky,"Effect of figure and waviness on image quality,"*Journal of optics* 14 (4), 189 (1983).

[113] S. Shikama,"Effects of corrugation of aspherical mirrors on the optical performance of imaging optics with the mirror near an image plane,"*Optical Engineering* 43 (12), 3068-3076 (2004).

[114] G.D. Boreman, *Modulation transfer function in optical and electro-optical systems*. Vol. 4. SPIE Press Bellingham, WA (2001).

[115] D.G. Voelz, *Computational fourier optics: a MATLAB tutorial*. Spie Press Bellingham, Wash, USA (2011).

[116] T.S. Hamidreza Aryan, Angela Davies, Christopher Evans, Miguel Alonso, Greg Forbes, MSF errors: effects and specifications for freefrom optics. University of North Carolina Charlotte (2016).

[117] V.N. Mahajan, "Optical Imaging and Aberrations, Part III: Wavefront Analysis." SPIE/International Society for Optical Engineering (2013).

[118] V.N. Mahajan,"Strehl ratio for primary aberrations: some analytical results for circular and annular pupils,"*JOSA* 72 (9), 1258-1266 (1982).

[119] J.E. Harvey,"Integrating optical fabrication and metrology into the optical design process,"*Applied optics* 54 (9), 2224-2233 (2015).

[120] F. Zernike,"The concept of degree of coherence and its application to optical problems,"*Physica* 5 (8), 785-795 (1938).

[121] J.B. Develis and G.B. Parrent Jr,"Transfer function for cascaded optical systems,"*JOSA* 57 (12), 1486-1487 (1967).

[122] J.E. Harvey, A. Krywonos, and D. Bogunovic,"Nonparaxial scalar treatment of sinusoidal phase gratings,"*JOSA A* 23 (4), 858-865 (2006).

[123] A. Krywonos, J.E. Harvey, and N. Choi,"Linear systems formulation of scattering

theory for rough surfaces with arbitrary incident and scattering angles,"*JOSA A* 28 (6), 1121-1138 (2011).

[124] J. Troutman, Z. Hosseinimakarem, Editor. (2017).

[125] C.-J. Kin and J. Wyant,"Subaperture test of a large flat or a fast aspheric surface,"*J. Opt. Soc. America, Vol. 71, p. 1587* 71, 1587 (1981).

[126] W.W. Chow and G.N. Lawrence,"Method for subaperture testing interferogram reduction,"*Optics letters* 8 (9), 468-470 (1983).

[127] T.W. Stuhlinger, "Subaperture optical testing: experimental verification." 1986 International Symposium/Innsbruck. International Society for Optics and Photonics p. 350-359 (1986).

[128] M. Chen, W. Cheng, and C.-W.W. Wang, "Multiaperture overlap-scanning technique for large-aperture test." San Diego-DL tentative. International Society for Optics and Photonics p. 626-635 (1992).

[129] W. Cheng and M. Chen,"Transformation and connection of subapertures in the multiaperture overlap-scanning technique for large optics tests,"*Optical Engineering* 32 (8), 1947-1950 (1993).

[130] P. Murphy, et al., "Subaperture stitching interferometry for testing mild aspheres." SPIE Optics+ Photonics. International Society for Optics and Photonics p. 62930J-62930J-10 (2006).

[131] M. Tricard, et al.,"Subaperture stitching interferometry of high-departure aspheres by incorporating a variable optical null,"*Cirp Annals-Manufacturing Technology* 59 (1), 547-550 (2010).

APPENDIX: MATLAB code

Contents

- Fit to Zernike Polynomials
- Organize coefficients in the coefficient map
- Sub functions

```
% Author: Zahra Hosseinimakarem
% Written July 2014, Revised Dec 2016
% This code gives Zernike fit coefficients of a NxN matrix
% using Jacobi polymomials in a recursive formula
% Inputs: Data map, Zernike order number
% Outputs: Zernike fit coefficients of the height map
%
_____INPUT_____________________________________________________
_____________
%   map                 raw data
%   nn1                 Zernike order number
%
%
_____OUTPUT____________________________________________________
_____________
%   Z      Zernike coefficients and plots of coefficients



function [Z]=ZpolyCoef(map,nn1);

% Crop valid data
idx=(~isnan(map));      % Makes a logical mask
[row col]=find(idx);    % Finds the rows and column of the
boundaries of mask
row_min=min(row(:));
row_max=max(row(:));
col_min=min(col(:));
col_max=max(col(:));
map_new=map(row_min:row_max, col_min:col_max);   % Crops the
data acroding to mask pixel numbers

map_rms=map_new(~isnan(map_new));
Ave=rms(map_rms(:)); % RMS of valid data
c_min=-3.*Ave;
c_max=3.*Ave;

map=map_new;
```

```
[nx ny]=size(map);
N=nx;
xs=1;
dx=2.*xs./N;  % Delta x in spatial domain
dy=dx;
X=linspace(-xs,xs,N);
Y=linspace(-xs,xs,N);
[x y]=meshgrid(X,Y);
r1=(x.^2+y.^2).^(0.5);

map(find(r1>0.99))=NaN;      % Replaces empty cells with NaN
figure;                    % Show input map
mesh(map);
xlabel('x(pixels)');
ylabel('y(pixels)');
title('Height map','fontsize',14,'fontweight','b');
shading interp;
axis xy;axis equal;
set(gcf,'color','w');
colorbar;
view([0 90]);
%% Fit to Zernike Polynomials
w=map;
nn=4;
arg=0.5.*((nn.*(nn+2)+nn))+1; % Number of terms
frame_tot=arg;

xs=1;        % Normalized aperture
dx=2.*xs./N;  % delta x in spatial domain
dy=dx;
X1=linspace(-xs,xs,N);
Y1=linspace(-xs,xs,N);
[x1 y1]=meshgrid(X1,Y1);
r=(x1.^2+y1.^2).^(0.5);
TH=-atan2(y1,x1);

rJ=(-N/2:1:(N/2-1))*dx;   % 1D Jacobi spatial domain
rJ=2.*rJ.^2-1;
frame_tot_2=0.5.*((nn1.*(nn1+2)+nn1))+1;
J2D=zeros(N,N);

Z=zeros(1,frame_tot_2);
w_temp=w(:);                  % linearize input map
w_temp=w_temp(~isnan(w_temp));   % select valid data
PL=zeros(frame_tot,N);          % The first set of Jacobis
from the function
PLM=zeros(frame_tot_2,N);       % More 1D Jacobi matrix
```

```
tic  % start the clock
j=1;
for n=0:nn;
    for m=(n):-2:-(n);
        mtemp=m;
        m=abs(m);
        if m==0;
            del=1;
        else
            del=0;
        end

        nor_f=((2.*(n+1))./(1+del)).^0.5;  % Normalization
factor
        numel=(n-m)./2;
        Result=Jacobi_abc(numel,n,m,rJ);
        PL(j,:)=Result;
        J1D=Result;

        for i=1:N
            J2D(i,:)=J1D(:);
        end

        J2D=interp2(X1,Y1,J2D,r,TH,'spline');
        Zrad=r.^(m).*J2D;

        if mtemp>=0
            Result=Zrad.*cos(m.*TH);
        else
            Result=Zrad.*sin(m.*TH);
        end

        polynm=Result;
        polynm=nor_f.*polynm;
        polynm=polynm(:);             % Linearize Zernike poly
        polynm=polynm(~isnan(w));
        Z(1,j)=polynm\w_temp;
        w_temp=w_temp-Z(1,j).*polynm;  % Subtract the fitted
Zernike from input map

        j=j+1;
    end
end

% The rest of Jacobis
```

```
k=1;
PLM(1:frame_tot,:)=PL;
for n=nn+1:nn1;
    for m=n:-2:-n;
        mtemp=m;
        m=abs(m);
        numel=(n-m)./2;
        numel_2=(n-2-m)./2;
        numel_3=(n-4-m)./2;
        if m==0;
            del=1;
        else
            del=0;
        end

        nor_f=((2.*(n+1))./(1+del)).^0.5;    % Normalization
factor

        if numel==0
            J1D=ones(1,N);
        elseif numel==1
            J1D=0.5.*(2+(m+2).*(rJ-1));
        else
            np=n;
            n2=np-2;
            n3=np-4;

            ind_1=0.5.*((np.*(np+2)+mtemp))+1;
            ind_2=0.5.*((n2.*(n2+2)+mtemp))+1;
            ind_3=0.5.*((n3.*(n3+2)+mtemp))+1;

            if numel_2==0
                P_2=ones(1,N);
            elseif numel_2==1
                P_2=0.5.*(2+(m+2).*(rJ-1));
            else
                P_2=PLM(ind_2,:);
            end

            if numel_3==0
                P_3=ones(1,N);
            elseif numel_3==1
                P_3=0.5.*(2+(m+2).*(rJ-1));
            else
                P_3=PLM(ind_3,:);
            end
            if (numel>1)
```

```
                J1D=(COZa(numel-1,m)+COZb(numel-
1,m).*rJ).*P_2-COZc(numel-1,m).*P_3;
                PLM(frame_tot+k,:)=J1D;
            end
        end

        % Converting 1D Jacobi to 2D Jacobi

        for i=1:N
            J2D(i,:)=J1D(:);
        end

        J2D=interp2(X1,Y1,J2D,r,TH,'spline');

        Zrad=r.^(m).*J2D;

        if mtemp>=0
            Result=Zrad.*cos(m.*TH);
        else
            Result=Zrad.*sin(m.*TH);
        end

        polynm=Result;
        polynm=nor_f.*polynm;
        polynm=polynm(:);
        polynm=polynm(~isnan(w));
        Z(1,k+frame_tot)=polynm\w_temp;
        w_temp=w_temp-Z(1,k+frame_tot).*polynm;
        k=k+1;
    end
end
toc

coeff_sample=Z;
save coeff_sample.mat  coeff_sample -mat -double % Saves
coefficients

%% Organize coefficients in the coefficient map
Z=Z';
k=1;
for n=0:nn;
    for m=(n):-2:-(n);
        j=0.5.*((n.*(n+2)+m));
        ZZ(k,1)=j;
        rad_in=ceil(0.5.*(-3+sqrt(9+8.*j)));
        rad(k,1)=rad_in;
        azi_in=2.*j-n.*(n+2);
```

```
        azi(k,1)=azi_in;
        k=k+1;
    end
end

X=azi;
Y=rad;
Z=abs(Z(1:arg,:));
x = [X(:); X(:); X(:)];
y = [Y(:); Y(:); Y(:)];
z = [Z(:); Z(:); Z(:)];

figure;
scatter3(x,y,z,[],z,'filled');colorbar;
view([0 90]);
xlabel('m','fontsize',12,'fontweight','b');
ylabel('n','fontsize',12,'fontweight','b');
title('Zernike fit
coefficients','fontsize',14,'fontweight','b');
set(gcf,'color','w');

np=linspace(0,nn,nn+1);
mp=linspace(0,nn,nn+1);
ZMp=NaN(nn+1,nn+1);

k=1;
for n=0:nn;
    for m=(n):-2:-(n);
        j=0.5.*((n.*(n+2)+m));
        ZZ(k,1)=j;
        azi_in=2.*j-n.*(n+2);
        if (azi_in >= 0)
            rad_in=ceil(0.5.*(-3+sqrt(9+8.*j)));
            ZMp(rad_in+1,azi_in+1)=Z(k);
        end
        k=k+1;
    end
end

ZMPS=ZMp.^2;              % Square of that map
ng=linspace(1,nn,nn);      %  n negative
mg=linspace(-nn,-1,nn);    %  m negative
ZMg=NaN(nn,nn);
k=1;
for n=0:nn;
    for m=(n):-2:-(n);
        j=0.5.*((n.*(n+2)+m));
```

```
        ZZ(k,1)=j;
        azi_in=2.*j-n.*(n+2);
        if (azi_in < 0)
            mneg=abs(azi_in);
            rad_in=ceil(0.5.*(-3+sqrt(9+8.*j)));
            ZMg(rad_in,mneg)=Z(k);
        end
        k=k+1;
    end
end

ZMgS=ZMg.^2;            % Square of neg map
nm=linspace(0,nn,nn+1);
mm=linspace(0,nn,nn+1);
mm_1=linspace(-2,nn,nn+3);
ZM=NaN(nn+1,nn+1);

% sum of pos and neg
ZM(:,1)=ZMPS(:,1);
ZMpp=ZMPS(2:nn+1,2:nn+1);
ZMp_neg=ZMgS+ZMpp;
ZM(2:nn+1,2:nn+1)=ZMp_neg;
ZM=sqrt(ZM);

log_ZM=log10(ZM);
log_ZM(find(log_ZM<(-5)))=NaN;  % Replaces very small values
with NaN
c_max=max(log_ZM(:));
c_min=-5;

ZM_big=NaN(nn+1,nn+3);
ZM_big(:,3:nn+3)=log_ZM;

img_data=ZM_big;

% Forbes indexing
F=NaN(nn./2+1,nn+1);

for n=0:nn
   for m=n:-2:0
        nf=(n-m)./2;
        F_dir(nf+1,m+1)=ZM(n+1,m+1);
        F(nf+1,m+1)=log_ZM(n+1,m+1);
    end
end

F_big=NaN(nn./2+1,nn+2);
```

```
F_big(:,2:nn+2)=F;

mf=linspace(0,nn,nn+1);
nf=linspace(0,nn./2,nn./2+1);

figure;imagesc(mf,nf,F);axis image;colorbar;
xlabel('m','fontsize',16,'fontweight','b');
ylabel('n_f','fontsize',16,'fontweight','b');
title('Fit coefficients, Log
scale','fontsize',14,'fontweight','b');
caxis([c_min, c_max]);
set(gcf,'color','w');
end

%% Subfunctions
% Subfunction, Recusrive generation of Jacobi polynomials
function Result=Jacobi_abc(numel,num,m,x);
% Recursive function to generate
% Zernike polynomials based on Jacobi polynomials
% References

% [1] Forbes, G. W. "Robust and fast computation for the
polynomials of optics."
%    Optics express 18.13 (2010): 13851-13862.
%    Based on Equations 1.6 and 4.1a,b,c
% [2] Abramowitz, M. and I.A. Stegun, Handbook of
mathematical functions. Applied mathematics series, 1966.
55: p. 62.

if numel==0
   Result=1;
elseif numel==1
   Result=0.5.*(2+(m+2).*(x-1));
else
   Result=(COZa(numel-1,m)+COZb(numel-
1,m).*x).*Jacobi_abc(numel-1,num,m,x)-COZc(numel-
1,m).*Jacobi_abc(numel-2,num,m,x);
end
end


% Subfunction, calculate the first recursive coefficient
function result = COZa(n,m)

% Based on M. Abramowitz, and I. Stegun, Handbook of
Mathematical Functions (Dover, 1978), Chap. 22
% eqn 22.7
```

```
a1=2.*(n+1).*(n+m+1).*(2.*n+m);
a2=-(2.*n+m+1).*m.^2;
a3=(2.*n+m).^3;
a4=2.*n.*(n+m).*(2.*n+m+2);
result=a2./a1;

end

% Subfunction, calculate the second recursive coefficient
function result = COZb(n,m)
% Based on M. Abramowitz, and I. Stegun, Handbook of
Mathematical Functions (Dover, 1978), Chap. 22
% eqn 22.7

a1=2.*(n+1).*(n+m+1);
a3=(2.*n+m+1).*(2.*n+m+2);
result=a3./a1;

end

% Subfunction, calculate the third recursive coefficient
function result = COZc(n,m)
% Based on M. Abramowitz, and I. Stegun, Handbook of
Mathematical Functions (Dover, 1978), Chap. 22
% eqn 22.7

a1=(n+1).*(n+m+1).*(2.*n+m);
a4=n.*(n+m).*(2.*n+m+2);
result=a4./a1;

end
```