OPTICAL SURFACE CHARACTERIZATION WITH THE STRUCTURE FUNCTION

by

Liangyu He

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
Mechanical Engineering

Charlotte

2013

Approved by:

____
Dr. Christopher J. Evans

____
Dr. Angela D. Davies

____
Dr. Edward P. Morse

____
Dr. Gregory J. Gbur

©2013
Liangyu He
ALL RIGHTS RESERVED

ABSTRACT

LIANGYU HE. Optical surface characterization with the structure function.
(Under the direction of DR. CHRISTOPHER J. EVANS and DR. ANGELA D. DAVIES)

It is important to characterize surface and transmitted wavefront errors in terms of the spatial content. The errors are typically analyzed in three spatial domains: figure, ripple (or mid-spatial frequency) and roughness. These errors can affect optical system performance. For example, mid-spatial frequency errors can lead to self-focusing and power loss in a high-power laser system. Currently, power spectral density (PSD) is used for the spatial content characterization in high-end optics, although there are potential pitfalls. For example, the low spatial content is removed before calculation, only a small fraction of surface data are used, and the results are sensitive to details like the windowing.

As an alternative, the structure function (SF) does not have such problems. It is the expectation of the squared height difference as a function of separation. The linear SF has been used in astronomy and captures data of all spatial frequencies. However, it does not capture anisotropy on the surface. The two-quadrant area SF introduced in this dissertation obviates this problem. It is computationally correct for any arbitrary aperture over all spatial content with anisotropic information.

This dissertation discusses some computational issues of the SF, which includes the calculation of the linear / area SF, sliding sampling method for large numbers of points within the aperture, analysis of periodic errors, and connection between the linear SF and area SF.

Moreover, the relationships between the SF and other surface characterization

techniques (Zernike polynomials, autocorrelation function (ACF), PSD, and RMS gradient) have been investigated. It turns out that the linear SF of the sum of the Zernike terms only equals to the sum of the linear SF of each of the Zernike polynomials with different azimuthal frequencies. However, this theorem does not apply to the area SF.

For stationary surfaces, the SF contains similar information as ACF, but it provides better visualization. The SF is computationally correct for any arbitrary aperture shape without extra processing, while the PSD always needs additional mathematical processing. After connecting the SF to the RMS gradient, the SF slope at the origin has been evaluated.

Use of a SF to specify optical surfaces over the full range of spatial frequencies of interest implies the combination of data from instruments with substantially different lateral resolutions. This research shows the combination of data from a Fizeau and a coherence scanning interferometer (CSI) for various precision surfaces. The investigation includes the connection method of the coordinate systems between the Fizeau data and the CSI sub-aperture data, the convergence of the averaged SF of sub-aperture samples, the uncertainty analysis, and the effect of the instrument transfer function (ITF).

In addition, the SF was used to explore two typical noise contributions (electronic noise and air turbulence) in phase shifting interferometry. Based on dynamic measurements, the SF was used to analyze the spatial components of a diamond turned surface after the compensation machining.

In summary, the SF is a useful tool to specify and characterize the spatial content of optical surfaces and wavefronts.

DEDICATION

To my beloved wife, Xiaoqin Xue, for her unconditional support and encouragement.

ACKNOWLEDGMENTS

Above all, I would like to express my deepest gratitude to my advisor, Dr. Christopher Evans, for his excellent guidance, caring, patience, and encouragement. Everytime I needed help, he always gave me the advice and support immediately even though he was really busy with his work. He made this research a pleasant experience for me, and without his extraordinary support I would never have been able to complete this project. It has been my great honor to work with him as his first doctoral student.

I would like to thank my co-advisor, Dr. Angela Davies, for her excellent guidance, patience, kindness and elegance. She has given me a lot of invaluable advice for my research. She also paid attention to details of my work, especially for the presentations for each Affiliate meeting and my defense. Her brilliant suggestions and great kindness made this project a challenging and joyful experience in my life.

I would like to thank Dr. Edward Morse, who has given me lots of help for my study at UNC Charlotte. I really appreciate everything he has done for me, especially for his invaluable reviews, comments and advice on my dissertation.

I would also like to thank Dr. Gregory Gbur for serving as my committee member and his time and effort spent in reviewing this dissertation.

I acknowledge Center for Precision Metrology (CPM) at UNC Charlotte for experimental support and the CPM Industrial Affiliates for funding this research, and my special thanks goes to Dr. Robert Hocken.

I would like to thank my friends, Kang Ni, Eric Browy, Mahsa Farsad, Yue Zhou, Benrui Zheng, Lucas Valdez, Mario Valdez and Mark Rubeo at CPM, UNC Charlotte. They always supported me and made my life enjoyable.

Finally, I would like to thank my parents for their understanding and support throughout my university career. My appreciation also goes to my beloved wife, Xiaoqin Xue. She was always there cheering me up and stood by me through the good times and bad.

TABLE OF CONTENTS

LIST OF TABLES xiii

LIST OF FIGURES xiv

LIST OF TERMS xxi

LIST OF ABBREVIATIONS xxii

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW 1

1.1 Motivation 1

1.2 Power Spectral Density (PSD) Function 4

1.2.1 Linear PSD 5

1.2.1.1 Definition of Linear PSD 5

1.2.1.2 Application of Linear PSD 7

1.2.2 Area PSD 8

1.2.2.1 Calculation of Area PSD 8

1.2.2.2 Application of Area PSD 10

1.3 Autocorrelation Function (ACF) 13

1.3.1 Linear ACF 13

1.3.1.1 Biased ACF 13

1.3.1.2 Unbiased ACF 15

1.3.1.3 Application of Linear ACF 16

1.3.2 Area ACF 16

1.3.2.1 Definition of Area ACF 16

1.3.2.2 Application of Area ACF 17

1.4 Structure Function (SF) 18

1.4.1 Linear Structure Function of Both Profile and Area Data 18

1.4.1.1 Astronomy 18

1.4.1.2 Rough Surfaces 20

1.4.1.3 General Optical Surfaces 24

1.4.2 Area Structure Function 26

1.5 Significance of This Project 27

1.6 Layout of Dissertation 28

CHAPTER 2: CALCULATION AND INTERPRETATION OF STRUCTURE FUNCTION 31

2.1 Linear Structure Function 31

2.1.1 Paired Sampling 32

2.1.2 Pure Sampling 33

2.1.3 Sliding Sampling 33

2.1.4 Comparison 33

2.2 Two-quadrant Area Structure Function 35

2.3 Relationship between Linear Structure Function and Area Structure Function 39

2.4 Spatial Frequency Interpretation with Structure Function 41

2.4.1 Simulated Periodic Errors 41

2.4.1.1 Sinusoidal Profile 41

2.4.1.2 Sinusoidal Surface 43

2.4.2 Diamond Turned Aluminum 44

2.5 Summary 47

CHAPTER 3: CHARACTERISTICS OF STRUCTURE FUNCTION AND RELATIONSHIP TO OTHER PARAMETERS 48

3.1 Structure Function and Zernike Polynomials 48

3.1.1 Zernike Polynomials 49

3.1.2 Linear SF of Zernike Polynomials 52

3.1.2.1 Linear SF of Individual Zernike Polynomial 53

3.1.2.2 Total Linear SF of a Surface Described by Zernike Polynomials 55

3.1.3 Area SF of Zernike Polynomials 61

3.1.3.1 Area SF of Individual Zernike Polynomial 61

3.1.3.2 Total Area SF of a Surface Comprised of Zernike Polynomials 63

3.2 Structure Function and Autocorrelation Function (ACF) 65

3.2.1 Linear SF and Linear ACF 65

3.2.2 Area SF and Area ACF 66

3.2.2.1 Stationary Surface 66

3.2.2.2 Non-stationary Surface 68

3.3 Structure Function and Power Spectral Density (PSD) Function 70

3.3.1 Linear SF and Linear PSD 70

3.3.1.1 Relationship between Linear SF and Linear PSD for Stationary Process 70

3.3.1.2 Comparison between Linear SF and Linear PSD 74

3.3.2 Area SF and Area PSD 85

3.3.2.1 Relationship between Area PSD and Area SF for Stationary Process 85

3.3.2.2 Comparison between Area SF and Area PSD 86

3.4 Root Mean Square (RMS) Gradient and Structure Function 90

3.5 Summary 92

CHAPTER 4: COMBINATION OF AREA STRUCTURE FUNCTIONS FROM MULTIPLE INSTRUMENTS 94

4.1 Instrument Transfer Function (ITF) 95

4.2 Challenge of Combination of SFs from Multiple Instruments 96

4.3 Convergence for Simulated Surfaces 101

4.3.1 Non-stationary Surface 101

4.3.2 Stationary Surface 105

4.3.3 Tilt Removal 107

4.4 Combination of Area Structure Functions 108

4.4.1 Polished Silicon Flat 109

4.4.1.1 Experimental Method 109

4.4.1.2 Uncertainty Associated with Tilt Correction 109

4.4.1.3 Area SF Convergence for CSI Measurements (20X) 112

4.4.1.4 Combination of Area SFs from Two Instruments 114

4.4.1.5 Type A Uncertainty in Combined Area SF 115

4.4.1.6 Combination of Area SFs without Tilt Correction 117

4.4.2 Diamond Turned Aluminum 118

4.4.2.1 SF Convergence for CSI Data (20X) 118

4.4.2.2 SF Combination from CSI (20X) and Fizeau Data 119

4.4.2.3 SF Combination from CSI (5X and 20X) and Fizeau Data 122

4.5 Summary 126

CHAPTER 5: APPLICATION OF STRUCTURE FUNCTION 127

5.1 Investigation of Electronic Noise and Air Turbulence with Structure Function 127

5.1.1 Uncertainty Analysis for Electronic Noise by Linear SF 128

5.1.2 Uncertainty Analysis for Air Turbulence by Linear SF 131

5.1.3 Uncertainty Analysis for Air Turbulence by Area SF 135

5.2 Dynamic Measurement on Diamond Turning Machine 137

5.3 Summary 142

CHAPTER 6: CONCLUSIONS AND FUTURE WORK 143

6.1 Conclusions 143

6.2 Future Work 145

6.2.1 Analysis of Type B Uncertainty for SF Combination 145

6.2.2 Extend the SF Combination from Flats to Spherical and Freeform Surfaces 146

6.2.3 Application of Two-quadrant Area SF in Astronomical Optics Specification 147

6.2.3.1 Linear and Area SF Specification of the TMT Segment 147

6.2.3.2 Simulation NO. 1 149

6.2.3.3 Simulation NO. 2 150

6.2.4 SF in General Optics Specification and Conformance Testing 152

REFERENCES 153

APPENDIX: MATLAB CODE 162

LIST OF TABLES

TABLE 2.1: Parameters of the linear SF in Figure 2.9. 42

TABLE 3.1: Zernike polynomials in Polar coordinates (the Fringe set (Loomis)). 50

LIST OF FIGURES

FIGURE 1.1: Deterministic polished surface analyzed in three different spatial frequency domains. 2

FIGURE 1.2: Effect of different spatial frequencies on the point spread function (PSF). 3

FIGURE 1.3:Transmitted wavefront spatial frequency specification for an NIF finished optic. 7

FIGURE 1.4: Area PSD of optical surface. 11

FIGURE 1.5: Radial PSD curves of each of the 3 sub-apertures in figure 1.4 (a), both with and without a Blackman window. 12

FIGURE 1.6: Calculation of biased ACF. 14

FIGURE 1.7: Calculation of unbiased ACF. 15

FIGURE 1.8: Relationship between ACF shape and grinding process. 16

FIGURE 1.9: Comparison of the area ACFs for two milled surfaces. 17

FIGURE 1.10: SF for Kolmogorov turbulence with a correlation length of 0.1 m and a wavelength of 0.5 μm. 19

FIGURE 1.11: Linear SF analysis for a GMT segment. 20

FIGURE 1.12: SF on log/log scale. 21

FIGURE 1.13: SF of a cylinder liner in worn and unworn conditions. 23

FIGURE 1.14: SF of surface profiles of limestone, sandstone, chert and carbonate. 24

FIGURE 1.15: Linear SF of the general optical surface. 25

FIGURE 1.16: The linear SF, RMS and RMS slope of the wavefront in figure 1.15 (a) and the respective specifications. 25

FIGURE 1.17: The linear SF and area SF of simulated map 1. 26

FIGURE 1.18: The linear SF and area SF of simulated map 2. 26

FIGURE 1.19: Three machined surfaces and their area SFs. 27

FIGURE 2.1: Calculation of linear SF of area data. 32

FIGURE 2.2: Sampling strategies. 32

FIGURE 2.3: Comparison of sampling strategies. 34

FIGURE 2.4: One-quadrant area SF. 35

FIGURE 2.5: Calculation of the area SF. 36

FIGURE 2.6: The area SFs in four quadrants. 39

FIGURE 2.7: Relationship between the linear SF and the area SF. 40

FIGURE 2.8: Relationship between the linear SF and the area SF. 40

FIGURE 2.9: Linear SF for 1D sinusoidal height data. 41

FIGURE 2.10: Linear SF for 1D combined sinusoidal height data. 42

FIGURE 2.11: SFs for 2D linear sinusoidal height data. 43

FIGURE 2.12: SFs for 2D radial sinusoidal height data. 44

FIGURE 2.13: Analysis for diamond turned aluminum surface. 44

FIGURE 2.14: Analysis for diamond turned aluminum surface. 45

FIGURE 2.15: Analysis for diamond turned aluminum surface. 46

FIGURE 3.1: Zernike polynomials ($n \leqslant 7$, $m \leqslant 7$). 53

FIGURE 3.2: Linear SFs of the Zernikes. 54

FIGURE 3.3: Sum of the two linear SFs of each of the Zernikes with different azimuthal frequencies. 57

FIGURE 3.4: Linear SF of the sum of Zernikes. 58

FIGURE 3.5: Difference between the SF of the total Zernikes and the sum of the two SFs. 58

FIGURE 3.6: Combination of the linear SFs with coefficients. 59

FIGURE 3.7: Sum of the two linear SFs of each of the Zernikes with the same azimuthal frequency. 60

FIGURE 3.8: Linear SF of the total Zernikes. 60

FIGURE 3.9: Difference between the SF of the total Zernikes and the sum of the two SFs. 61

FIGURE 3.10: Area SF of astigmatism. 61

FIGURE 3.11: Area SF of coma. 62

FIGURE 3.12: Area SF of trefoil. 62

FIGURE 3.13: Area SF of defocus. 63

FIGURE 3.14: Area SF of third order spherical. 63

FIGURE 3.15: Sum of two area SFs of each of the Zernike polynomials with different azimuthal frequencies. 64

FIGURE 3.16: Area SF of the total Zernikes. 64

FIGURE 3.17: Difference between the SF of the total Zernikes and the sum of the two SFs. 65

FIGURE 3.18: Zernikes fit of a polished surface. 67

FIGURE 3.19: Relationship between area SF and area ACF for a polished surface after removing 36 Zernikes. 68

FIGURE 3.20: Relationship between area SF and area ACF for a non-stationary surface. 68

FIGURE 3.21: Analysis of a polished-silicon flat. 69

FIGURE 3.22: Comparison of PSD based on profile data and ACV. 72

FIGURE 3.23: Calculation of the linear PSD based on the SF. 74

FIGURE 3.24: Simulated sinusoidal surface with λ= 400 µm. 75

FIGURE 3.25: Calculation of linear PSD without window function. 76

FIGURE 3.26: Linear SF of the sinusoidal surface. 77

FIGURE 3.27: Hanning and Hamming windows in spatial domain. 79

FIGURE 3.28: Calculation of linear PSD in different ways. 80

FIGURE 3.29: Combination of two surfaces. 82

FIGURE 3.30: Calculation of linear PSD with Hanning window. 83

FIGURE 3.31: Linear SF of the combined surface. 84

FIGURE 3.32: Linear SF of the combined profile. 85

FIGURE 3.33: Analysis of the use of window. 87

FIGURE 3.34: Analysis of a diamond-turned aluminum flat. 88

FIGURE 3.35: Analysis of simulation data. 89

FIGURE 3.36: Analysis of simulation data. 90

FIGURE 4.1: The theoretical (dashed lines) and experimental (solid lines) ITF for four objectives used in the interferometric microscope. 96

FIGURE 4.2: Measurement with Zygo Verifire AT Fizeau interferometer. 97

FIGURE 4.3: Measurement with Zygo ZeGage CSI. 97

FIGURE 4.4: Silicon surface measured with different instruments. 98

FIGURE 4.5: Some parameters of the two sub-apertures. 98

FIGURE 4.6: Area SFs of the two sub-apertures. 99

FIGURE 4.7: Diamond turned aluminum measured with different instruments. 99

FIGURE 4.8: Some parameters of the two sub-apertures. 100

FIGURE 4.9: Area SFs of the two sub-apertures. 100

FIGURE 4.10: Sampling method. 102

FIGURE 4.11: Area SF convergence for simulated radial sinusoidal surface. 103

FIGURE 4.12: Area SF convergence indicates a dependence consistent with $(N)^{-1/2}$ for simulated sinusoidal surface. 104

FIGURE 4.13: Simulated random noise. 105

FIGURE 4.14: Area SF convergence for simulated random noise. 106

FIGURE 4.15: Area SF convergence indicates a dependence consistent with $(N)^{-1/2}$ for simulated random noise. 106

FIGURE 4.16: Tilt analysis of a simulated surface. 108

FIGURE 4.17: Polished silicon flat. 110

FIGURE 4.18: Curvature. 112

FIGURE 4.19: Flow chart of calculating RMS of ΔSF. 113

FIGURE 4.20: Area SF convergence for polished Si flat (CSI 20x objective). 114

FIGURE 4.21: Combination of area SF based on the Si flat measured with the Fizeau interferometer and the CSI (20x). 115

FIGURE 4.22: Variation for a profile from the combined area SF. 116

FIGURE 4.23: Combination of area SF based on the Si flat measured with Fizeau interferometer and CSI 20x (no tilt-correction). 117

FIGURE 4.24: Variation for a profile from the combined area SF (no tilt-correction). 117

FIGURE 4.25: Diamond turned Al. 118

FIGURE 4.26: Area SF convergence for diamond turned Al and the comparison with polished Si flat (CSI 20x). 119

FIGURE 4.27: Combination of area SF based on the diamond turned Al measured with Fizeau interferometer and CSI (20x objective). 120

FIGURE 4.28: Variation for a profile from the combined area SF. 120

FIGURE 4.29: Combination of area SF based on the diamond turned Al measured with Fizeau interferometer and CSI (20x objective). 121

FIGURE 4.30: Variation for a profile from the combined area SF. 122

FIGURE 4.31: Combination of area SF based on the diamond turned Al measured with Fizeau interferometer and CSI (5x objective). 123

FIGURE 4.32: Variation for a profile from the combined area SF. 124

FIGURE 4.33: Combination of the SF for Fizeau data and CSI data (at 5x and 20x). 125

FIGURE 5.1: Laser Fizeau interferometer configuration. 127

FIGURE 5.2: Setup of the interferometer with a short insulated cavity. 128

FIGURE 5.3: Measurement with Fizeau interferometer. 129

FIGURE 5.4: SF of the difference. 130

FIGURE 5.5: SF analysis for random noise. 131

FIGURE 5.6: Setup of the interferometer with different cavity lengths. 131

FIGURE 5.7: SFs of the differences. 133

FIGURE 5.8: Turbulence parallel distribution (introducing piston). 134

FIGURE 5.9: Turbulence vertical distribution (introducing tilt). 134

FIGURE 5.10: Turbulence radial distribution (introducing power). 134

FIGURE 5.11: Linear SF of the power. 135

FIGURE 5.12: Zygo DynaFiz interferometer. 135

FIGURE 5.13: Setup of the DynaFiz interferometer. 136

FIGURE 5.14: Area SF of turbulence. 137

FIGURE 5.15: Measurement configuration. 138

FIGURE 5.16: Experiment setup after the first machining. 139

FIGURE 5.17: Measured height maps. 139

FIGURE 5.18: Measured height maps. 140

FIGURE 5.19: Area SF in four quadrants. 141

FIGURE 6.1: SF specification of TMT segment surface FIGURE accuracy. 148

FIGURE 6.2: SF of the simulated surface error. 149

FIGURE 6.3: SF analysis with simulated surface error. 150

FIGURE 6.4: SF of the simulated surface. 150

FIGURE 6.5: SF analysis with simulated surface error. 151

LIST OF TERMS

Autocovariance function (ACV) - It is the integral of the shifted and unshifted profiles/surfaces evaluated over the length/area of the profile/surface.

Autocorrelation function (ACF) - It is the ACV normalized by the square of the root mean square roughness.

Instrument transfer function (ITF) - In surface metrology, a typical ITF is the optical transfer function (OTF), which describes the optical system's ability to reproduce images in terms of spatial frequency (inverse length). The modulus of the OTF is the well-known modulation transfer function (MTF).

Power spectral density function (PSD) - It is the square of the Fourier transform of the profile/surface, appropriately normalized. It is based on the spatial frequency in inverse length units.

Structure function (SF) - It is the expectation of the squared height difference as a function of separation.

Total structure function - It is the structure function of a surface described by a set of Zernike polynomials in this dissertation.

LIST OF ABBREVIATIONS

ACF	autocorrelation function
ACV	autocovariance function
AFM	atomic force microscope
CSI	coherence scanning interferometer
DFT	discrete Fourier transform
DTM	diamond turning machine
FFT	fast Fourier transform
ITF	instrument transfer function
MRF	magnetorheological finishing
MTF	modulation transfer function
NIF	national ignition facility
OTF	optical transfer function
PSD	power spectral density function
PSF	point spread function
PSI	phase shifting interferometer
PV	peak to valley
PVr	robust peak-to-valley specification
Ra	roughness average
RMS	root mean square
SF	structure function
SWLI	scanning white light interferometer
TMT	thirty meter telescope

CHAPTER 1 : INTRODUCTION AND LITERATURE REVIEW

1.1 Motivation

Engineers and scientists have expended significant energy in recent decades trying to close the loop between surface specification, measurement, and functional performance (to predict tribological and heat transfer behaviors of surface contacts) [1,2]. Generally, the easily calculated amplitude parameters such as Ra or PV are insufficient, and a more complex characterization of the surface is required [3]. Methods such as fractal and scale-sensitive analysis [4], spatial functions of areal data [5], slopes, and asperity radius [6,7] have been applied in different circumstances. This evolution was motivated by the complementary needs of manufacturing process control and the surface topography connection to performance.

In optics, in particular, it is often important to characterize surface and transmitted wavefront errors in terms of the spatial content. For traditional optical fabrication, the similar geometrical size between tool and optical part can generate long spatial scale surface undulations (figure or form errors), and the surface fracture mechanics can introduce short spatial scale undulations (roughness or scatter) [8]. In this case, optical surfaces can be specified by PV or RMS. More recently, a robust amplitude parameter, PVr, which combines the PV of a 36-term Zernike fit and RMS of the residual, has been introduced by Christopher. J. Evans [9,10]. However, For modern optical fabrication, the deterministic techniques such as diamond turning, magnetorheological finishing (MRF)

and computer-controlled polishing can easily generate mid-spatial scale surface undulations (ripples), which accumulate locally where the removal of material departs from the specified values and could lead to small-angle scatter that degrades optical performance [11,12].

Therefore, in modern optical fabrication, the surface errors are typically analyzed in three spatial domains [13,14,15]: figure (low spatial frequency), ripple (mid-spatial frequency [16,17,18,19,20]) and roughness (high spatial frequency), often with arbitrary cut-offs. (The choice of cut-offs can be derived from the aperture and function of the optical component. However, the available standards take a different approach. See ISO 10110 and 25178.) These errors can affect optical system performance. For example, mid-spatial frequency errors can lead to self-focusing and power loss in a high-power laser system.

Figure 1.1 shows a surface generated with a deterministic polishing method decomposed into three different spatial frequency domains. In the mid-spatial frequency component, the unidirectional scanning sub-aperture tool path (raster polishing) is clear [8, 21].

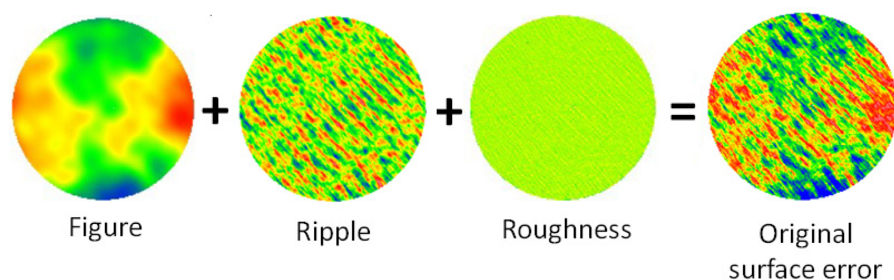


Figure 1.1 Deterministic polished surface analyzed in three different spatial frequency domains [8,21].

Currently, the power spectral density function (PSD) is used for analyzing the surface spatial frequency components [22] in high end optics. It is plotted in units of length cubed versus spatial frequency in inverse length units, and its relationship to functional performance was graphically captured by Harvey and Kotha [23], showing the optical consequences of different spatial frequencies (Figure 1.2).

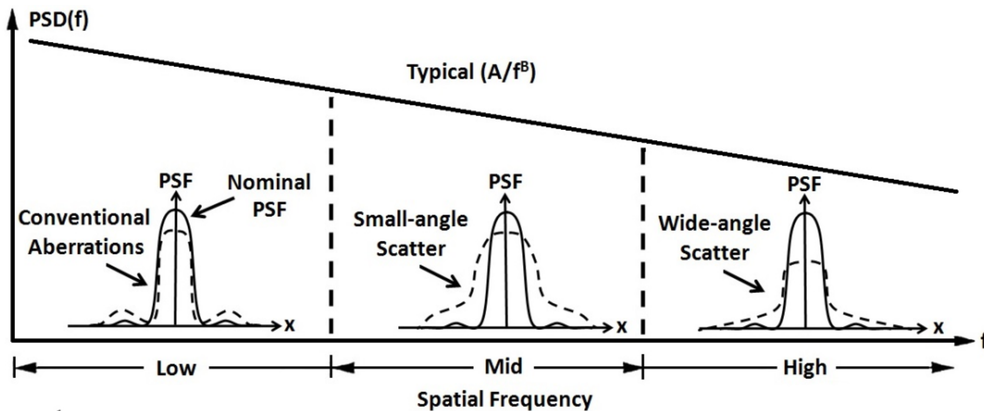


Figure 1.2 Effect of different spatial frequencies on the point spread function (PSF) (after [23]).

The PSD has been used in characterization and specification [24] of X-ray optics, neutron and other advanced optics, as well as in the characterization of machined surfaces. PSDs, however, cannot be used to characterize all components of the surface error; errors with surface wavelengths of 3 cycles/aperture or longer must be treated separately. Only profiles or rectangular areas can be analyzed. In addition, the reported PSD is sensitive to computational details such as windowing and zero padding. This will be discussed in more detail in Chapter 3.

Less often, the auto-correlation function (ACF) has also been used to characterize the spatial content of surfaces, but it has a number of limitations. As the calculation of ACF is based on the product of two amplitudes, it is strongly dependent on the mean plane and is

not intuitive. Moreover, the ACF cannot describe well a non-stationary mean, which is a common feature of engineering or optical surfaces [25].

The structure function (SF) – the squared expectation of height difference as a function of separation – has some advantages compared with the other spatial representations. The computational details will be given in Chapter 2 along with some examples. Thus, it is an alternative of the PSD and ACF. The relationship among SF, PSD and ACF will be discussed in Chapter 3. The SF has been used in surface metrology since the 1970s [25, 26], although not extensively, and in astronomy to specify the surface of large optics. A linear SF can be computed for profile or area data, with the loss of anisotropic detail. The area SF can be calculated without filtering, zero padding or restraints on aperture size or shape. Also, it retains information on anisotropy [27,28,29].

This chapter starts with definitions, applications and limitations of the PSD and ACF in surface metrology. Next, the definition and current applications of SF will be described. Finally, the purpose and significance of this project will be discussed.

1.2 Power Spectral Density (PSD) Function

Basically, PSD is a mathematical concept based on the Fourier transform. It describes power of the variations as a function of spatial frequency. Conventionally, PSD has wide application in the electronic and mechanical engineering in signal analysis [30] and vibration estimation [31] in the frequency domain. Recently, it has been used for the optical characterization [32,33,34,35], especially for quantifying the spectrum of surface and wavefront quality in optical systems [22]. A typical application is the optical specifications of mid-spatial frequency analysis in the National Ignition Facility

[11,12,36,37] at Lawrence Livermore National Laboratory for which thousands of large optical components have been fabricated.

1.2.1 Linear PSD

1.2.1.1 Definition of Linear PSD

(1) One-side linear PSD

For a discrete profile, the PSD can be calculated based on the discrete Fourier transform (DFT):

$$PSD(f_m) = \frac{2\Delta x}{N} \left| DFT(f_m) \right|^2 K(m), \qquad (0 \le m \le \frac{N}{2}) \tag{1.1}$$

where DFT is usually evaluated with the fast Fourier transform (FFT) algorithm, which greatly increases the computational speed. Thus, the PSD can be expressed as:

$$PSD(f_m) = \frac{2\Delta x}{N} \left| \sum_{n=0}^{N-1} e^{-i2\pi mn/N} z(n) \right|^2 K(m), \qquad (0 \le m \le \frac{N}{2}) \tag{1.2}$$

where Δx is the sampling interval of the profile measurements, N is the total number of sampling points, n is the sampling index, $z(n)$ is the profile height data, and f_m is the spatial frequency (reciprocal of the spatial wavelength) described as $f_m = \frac{m}{N\Delta x}$ (m ranges from 0 to $N/2$).

When $m = N/2$, the sampling frequency $f_m = \frac{1}{2\Delta x}$. This frequency is the Nyquist frequency of the sampled measurement [38].

The factor of 2 rises from the efficient FFT algorithm that only calculates the results for positive frequencies. In order to satify Parseval's theorem, the factor 2 should be applied. Moreover, $K(m)$ is a bookkeeping factor: $K(m) = 1$ except $K(0) = K(N/2) = 1/2$. It also ensures that the calculation is consistent with Parseval's theorem:

$$\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2, \qquad (1.3)$$

where $X(k)$ is the DFT of $x(n)$, both of length N.

In this case, Parseval's theorem is expressed in the form as:

$$\begin{aligned} RMS^2 &= \frac{1}{N} \sum_{n=0}^{N-1} |z(n\Delta x)|^2 = \frac{1}{N^2} \sum_{k=0}^{N-1} |DFT(k)|^2 = \frac{2}{N^2} \sum_{m=0}^{N/2} \left| \sum_{n=0}^{N-1} e^{-i2\pi mn/N} z(n\Delta x) \right|^2 K(m) \\ &= \frac{1}{N\Delta x} \sum_{m=0}^{N/2} PSD(f_m) = f_1 \sum_{m=0}^{N/2} PSD(f_m). \end{aligned} \qquad (1.4)$$

In other words, the "area" under the PSD curve in any frequency interval is exactly the square of the RMS of the profile over the same frequency interval [12].

(2) Two-side linear PSD

The linear PSD can also be defined on two sides: [32]

$$\begin{aligned} PSD(f_m) &= \frac{\Delta x}{N} |DFT(f_m)|^2 K(m) \\ &= \frac{\Delta x}{N} \left| \sum_{n=0}^{N-1} e^{-i2\pi mn/N} z(n) \right|^2 K(m), \qquad (-\frac{N}{2} \le m \le \frac{N}{2}) \end{aligned} \qquad (1.5)$$

where both results for positive and negative frequencies are considered and the factor of 2 in the one-side linear PSD is not necessary. The *K(m)* is a bookkeeping factor: *K(m)* = 1 except *K(±N/2)* = 1/2. When plotting the PSD curve, *m* usually ranges from 1 to *N*/2, which means the spatial frequency ranges from $f_1 = \frac{1}{N\Delta x}$ to $f_{N/2} = \frac{1}{2\Delta x}$. However, when considering Parseval's theorem, *m* should range from -*N*/2 to *N*/2, that is

$$\begin{aligned} RMS^2 &= \frac{1}{N} \sum_{n=0}^{N-1} |z(n)|^2 = \frac{1}{N^2} \sum_{k=0}^{N-1} |DFT(k)|^2 = \frac{1}{N^2} \sum_{m=-N/2}^{N/2} \left| \sum_{n=0}^{N-1} e^{-i2\pi mn/N} z(n) \right|^2 K(m) \\ &= \frac{1}{N\Delta x} \sum_{m=-N/2}^{N/2} PSD(f_m) = f_1 \sum_{m=-N/2}^{N/2} PSD(f_m). \end{aligned} \qquad (1.6)$$

These two definitions are the same in theory - they simply have different representations. The biggest difference is about the range of the f_m. For the first definition, $0 \le m \le \frac{N}{2}$; while for the second definition, $-\frac{N}{2} \le m \le \frac{N}{2}$. In other words, the first definition only contains zero and positive spatial frequencies, while the second definition contains all the possible frequencies ranging from negative to positive. In fact, the PSD of the negative frequencies are identical to that of the corresponding positive frequencies.

1.2.1.2 Application of Linear PSD

Here is an example of the linear PSD used for one of the national ignition facility (NIF) optics (J. H. Campbell et at, 2004).

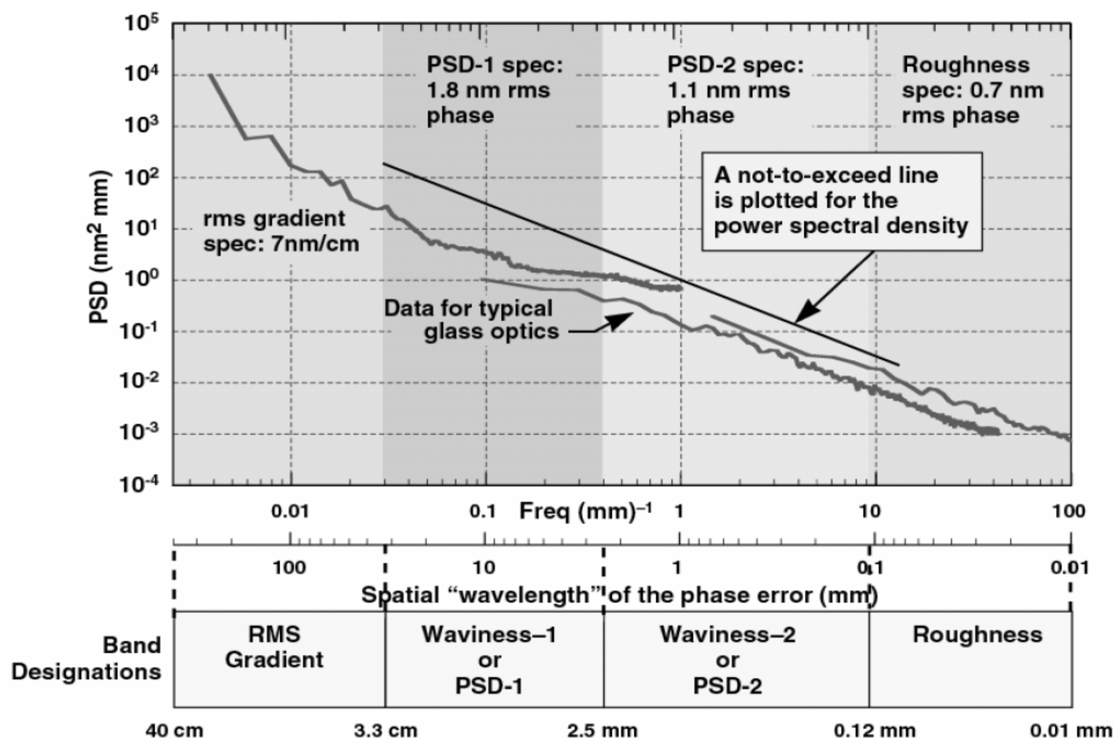


Figure 1.3 Transmitted wavefront spatial frequency specification for an NIF finished optic [39].

As Figure 1.3 shows, the oblique straight line is the maximum allowable PSD for spatial frequencies from 33 mm to 0.12 mm (0.03 mm^{-1} to 8.5 mm^{-1}). In other words, the PSD of the fabricated optics should not exceed this line. The other parameter for this specification is the RMS in terms of specific frequency ranges, that is, the area beneath the PSD curve with related spatial frequencies. For example, the RMS over the frequency domain from 1/33 mm^{-1} to 1/2.5 mm^{-1} (i.e. the area beneath the PSD-1 specification) should be less than the 1.8 nm required [12,39].

Thus, the PSD specification is a good way to control spatial frequency errors on a transmitted wavefront.

1.2.2 Area PSD

With the development of 3-D surface measuring instruments such as the scanning white light interferometer (SWLI), the phase shifting interferometer (PSI) and the atomic force microscope (AFM), researchers have extended the linear analysis to areal investigations [40] in order to obtain a better and comprehensive representation of the surface.

1.2.2.1 Calculation of Area PSD

(1) The definition of the area PSD

Similar to the linear PSD, the calculation of area PSD is based on the 2 dimensional discrete Fourier transform (DFT):

$$\begin{aligned} PSD(f_x, f_y) &= \frac{\Delta x \Delta y}{N_x N_y} \left| DFT(f_x, f_y) \right|^2 \\ &= \frac{\Delta x \Delta y}{N_x N_y} \left| \sum_{n_x=0}^{N_x-1} \sum_{n_y=0}^{N_y-1} e^{-i2\pi\left(f_x n_x \Delta x + f_y n_y \Delta y\right)} z(n_x \Delta x, n_y \Delta y) \right|^2, \end{aligned} \qquad (1.7)$$

where Δx and Δy are the sampling intervals of the surface measurements in x and y, N_x and N_y are the total number of sampling points, n_x and n_y are the sampling indices, $z(n_x\Delta x, n_y\Delta y)$ are the surface height data, and f_x and f_y are the spatial frequences (reciprocal of the spatial wavelength) described as $f_x = \frac{m}{N_x\Delta x}$ (m ranges from -N_x/2 to N_x/2) and $f_y = \frac{n}{N_y\Delta y}$ (n ranges from -N_y/2 to N_y/2), respectively.

When $m = N_x/2$, the sampling frequency $f_x = \frac{1}{2\Delta x}$. This frequency is the Nyquist frequency of the sampled measurement in x direction. So is f_y.

In fact, each PSD of the negative frequency is identical to that of the corresponding positive frequency.

(2) Relationship between RMS^2 and the area PSD

For better demonstration, let $u = f_x$ and $v = f_y$, then

$$PSD(u_i, v_j). \quad (0 \le i \le N_x - 1;\ 0 \le j \le N_y - 1) \tag{1.8}$$

According to Parseval's theorem,

$$\begin{aligned} RMS^2 &= \frac{1}{N_x N_y}\sum_{m=0}^{N_x-1}\sum_{n=0}^{N_y-1}\left|z(x_m, y_n)\right|^2 = \frac{1}{N_x N_y}\left\{\frac{1}{N_x N_y}\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1}\left|DFT(u_i, v_j)\right|^2\right\} \\ &= \frac{1}{N_x\Delta x\, N_y\Delta y}\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} PSD(u_i, v_j). \\ &= \Delta u\, \Delta v \sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} PSD(u_i, v_j) \end{aligned} \tag{1.9}$$

In other words, the "volume" under the PSD curved surface in any frequency interval is exactly the square of the RMS of the measured surface over the same frequency interval [41].

(3) Relationship between the linear PSD and area PSD

After integration [11,22], the area PSD can be transformed to the linear PSD:

$$PSD(u_i) = \frac{1}{\Delta y}\sum_{j=0}^{N_y-1} PSD(u_i, v_j) = \Delta v \sum_{j=0}^{N_y-1} PSD(u_i, v_j). \quad (0 \le i \le N_x - 1) \qquad (1.10)$$

Eq. (1.10) shows the linear PSD in x direction.

Note that the units of the linear and area PSD are different. For the linear PSD, it is length cubed; for the area PSD, it is length to the fourth power.

If the surface is isotropic, the $PSD(f_x, f_y)$ can be expressed in polar coordinates [34] $PSD(f, \theta)$, where

$$\begin{cases} f = \sqrt{f_x^{\,2} + f_y^{\,2}} \\ \theta = \tan^{-1}\left(\dfrac{f_y}{f_x}\right) \end{cases}. \qquad (1.11)$$

Then the linear PSD can be written

$$PSD(f) = f \int_0^{2\pi} PSD(f, \theta) d\theta. \qquad (1.12)$$

1.2.2.2 Application of Area PSD

The calculation of PSD is based on the discrete Fourier transform (DFT) which always processes a profile/image as if it were part of a periodic function of identical curves/images extending to infinity. Thus, there are strong edge effects between the neighbors of such a periodic function. To solve this problem, it is important to multiply the data with a “window function” before applying the DFT. The window function starts around zero, then increases to a maximum at the center of the curve/image and decreases again [42]. Thus the impact of the edge discontinuity can be minimized.

Here is an example given by P. Z. Takacs using the area PSD. As Figure 1.4 (a) shows, P. Z. Takacs chose three representative sub-apertures from the wavefront map of a lens, detrended the low order forms, and calculated their area PSDs with a Blackman window. For example, Figure 1.4 (b) is the windowed area PSD of residual error from the rectangular sub-aperture below center in Figure 1.4 (a) after detrending by removal of a 4th order polynomial. It is obvious that there are some directional components.

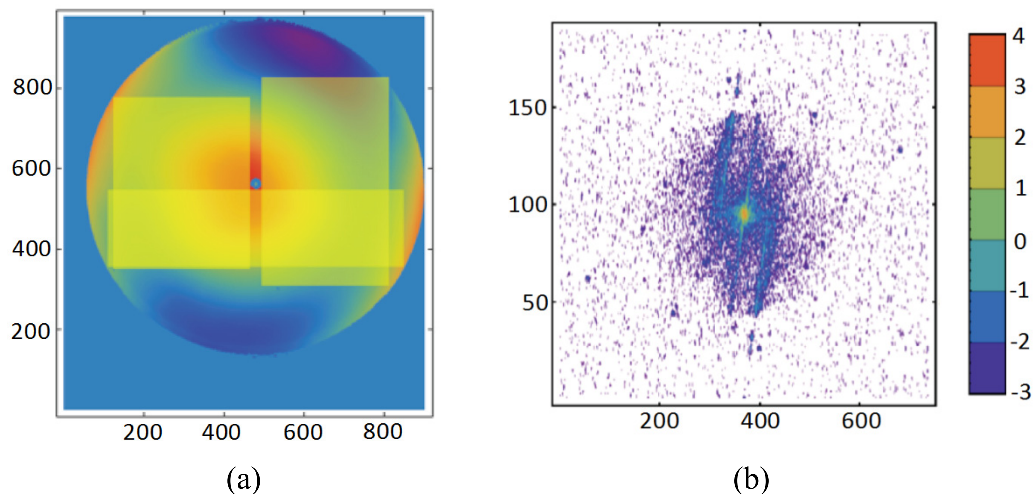


Figure 1.4 Area PSD of optical surface [43]. (a) Wavefront map of a lens measured with a Fizeau interferometer. (b) Windowed area PSD of residual error from the rectangular sub-aperture below center in Figure 1.4 (a) after detrending 4th order polynomial. (Vertical color scale is log PSD in units of μm^4.)

To make extracting surface statistics easier, P. Z. Takacs integrated the area PSD over all azimuthal angles and generated a radial PSD. Figure 1.5 shows 2 sets of 3 radial PSD curves (corresponding to 3 sub-apertures in Figure 1.4 (a)), both with and without a Blackman window. In addition, the average of the 3 curves for each set is shown in heavy lines.

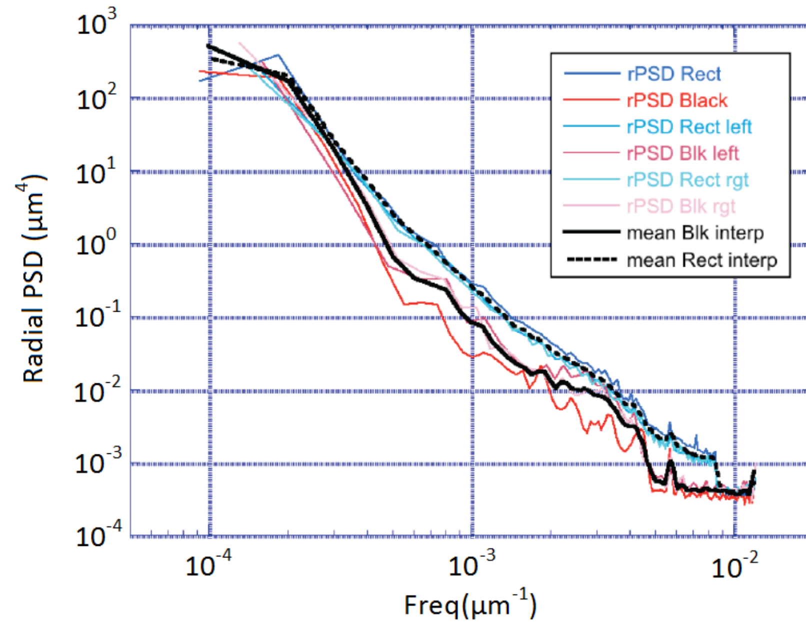


Figure 1.5 Radial PSD curves of each of the 3 sub-apertures in Figure 1.4 (a), both with and without a Blackman window [43].

In this case, the RMS of mid-spatial frequency components from 5 mm to 100 μm is specified, which can be calculated from the area under the PSD curve over the specified bandwidth. The specification of the RMS for this surface is 0.5 nm. Based on the averaged curves, the related RMS is 0.38 nm for the Blackman windowed data and 0.74 nm for the non-windowed data. In order words, with the Blackman window, the RMS is within the tolerance and the lens passes; without the window, the RMS exceeds the specification and the lens fails. Thus, improper data processing could cause economic loss if a good part is rejected [43], assuming that the use of a Blackman window meets the designers intent.

In summary, the PSD is a good tool for specifying optical surfaces, but it is sensitive to data processing such as windowing and sub-aperture choosing. Different calculation techniques may lead to different results.

1.3 Autocorrelation Function (ACF)

A breakthrough in surface characterization was the use of stochastic techniques, for example, the ACF. It has been an important tool to characterize rough surfaces, especially for representing spatial variation (the rapidity with which the height varies with distance in a certain direction).

The ACF was first used in surface metrology by J. R. Womersley and M. R. Hopkins in 1945 [44], but it was J. Peklenik [45] who used the ACF in terms of a typology for surfaces. Since then, the characterization of surfaces with the statistical analysis of height data has become more acceptable than the distribution of peaks and valleys [46]. However, the discrete analysis of the ACF was originated by D. J. Whitehouse [47] who, with J. F. Archard derived a functional significance from the distance over which it decays.

1.3.1 Linear ACF

The concept of correlation is well known in statistics, where it is used to analyze the influence between two sets of numbers. When calculating the linear ACF, the original profile should be duplicated, and the duplicated profile is shifted along the original profile with a separation τ, then the product of the two profiles in the overlapped region is obtained. By repeating this process, the whole linear ACF can be generated for every separation τ.

1.3.1.1 Biased ACF

The biased autocovariance function (ACV) is given as

$$ACV(\tau) = \frac{1}{n}\sum_{i=1}^{n-\tau} z(i)z(i+\tau). \quad 0 \le \tau \le n-1 \qquad (1.13)$$

The ACF is the surface height covariance normalized to the mean square σ^2, which is defined as

$$ACF(\tau) = \frac{ACV(\tau)}{\sigma^2}, \tag{1.14}$$

where

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} z(i)z(i). \tag{1.15}$$

Figure 1.6 (a) is a simulated sinusoidal function with λ = 20 mm, and Figure 1.6 (b) is its biased ACF.

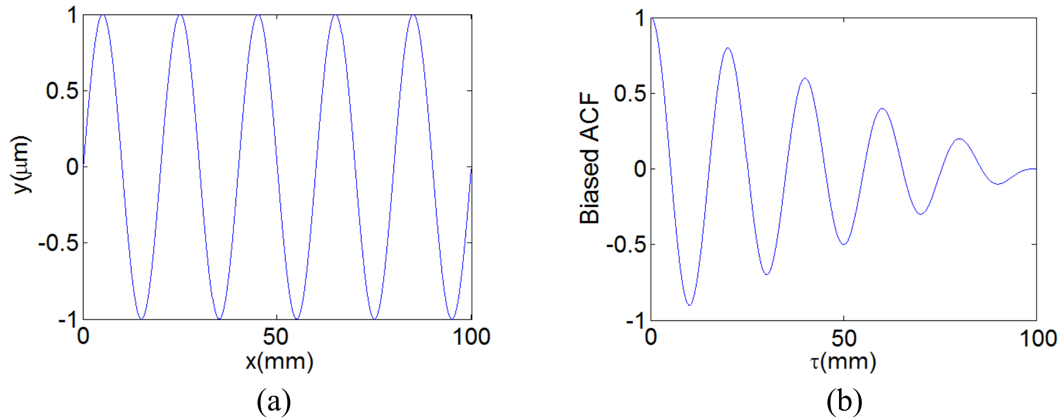


Figure 1.6 Calculation of biased ACF. (a) Sinusoidal function. (b) Biased ACF.

An important parameter of the ACF is the correlation length. It is the length of a profile over which the ACF drops to a small fraction of its original value. In this case, as the profile is a periodic function, the correlation length can be defined as the length over which the ACF drops below the threshold value of zero [48], which is 5 mm here.

The other important property of the ACF can be observed in Figure 1.6 (b) is the periodicity [48]. Although the amplitude of the ACF is linearly damped due to the calculation with smaller overlapped region and constant averaging number *n*, there are periodic peaks and valleys. When the duplicated profile is out of phase with the original profile by $2k\pi$, where *k* is an integer, the ACF shows the peaks. When they are out of

phase by $(2k+1)\pi$, the ACF shows the valleys. In brief, the ACF reflects the same period as the input profile (20 mm).

1.3.1.2 Unbiased ACF

The unbiased ACV is given as:

$$ACV(\tau)=\frac{1}{n-\tau}\sum_{i=1}^{n-\tau}z(i)z(i+\tau), \quad 0\le\tau\le n-1. \tag{1.16}$$

Unless otherwise stated, we will use the unbiased ACV and ACF for the discussion in this dissertation.

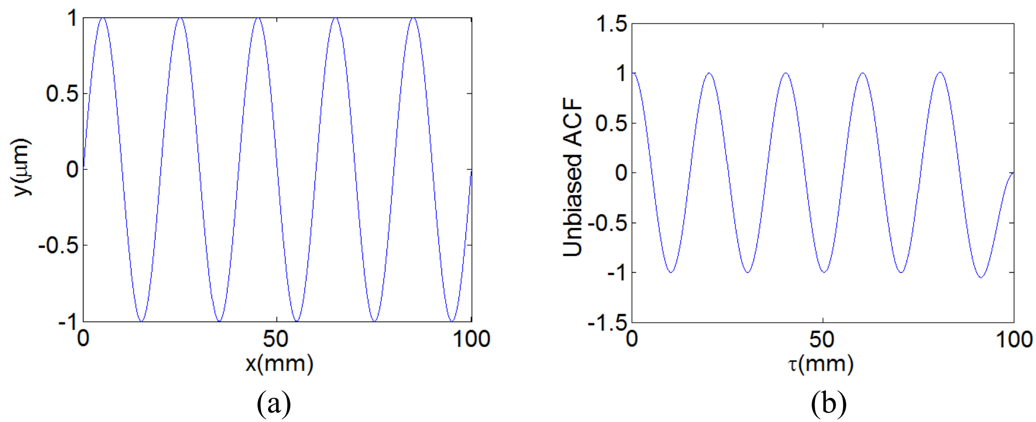


Figure 1.7 Calculation of unbiased ACF. (a) Sinusoidal function. (b) Unbiased ACF.

Figure 1.7 shows the calculation of unbiased ACF for the same profile in Figure 1.6 (a). The biggest difference is that the unbiased ACF does not taper off, because the averaging number $(n-\tau)$ exactly reflects the overlapped region. The large amplitude of the last valley stems from the calculation with only a few points from the profile, which can not reflect the real characteristics. In other words, the ACF with shorter separations is more accurate due to the calculation with more points.

Under ideal conditions, if the length of the input profile were infinite, neither the biased and unbiased ACFs would taper off.

1.3.1.3 Application of Linear ACF

D. J. Whitehouse has applied the linear ACF to analyze mechanically ground surfaces [46]. Figure 1.8 shows the relationship between the ACF shape and the grinding process. For example, the randomness of the process (how the grains are distributed in the wheel surface) determines the envelope of the ACF of the surface texture. Moreover, the shape of the grains and how each grain interacts with the surface will determine the fine detail of the ACF as shown in Figure 1.8.

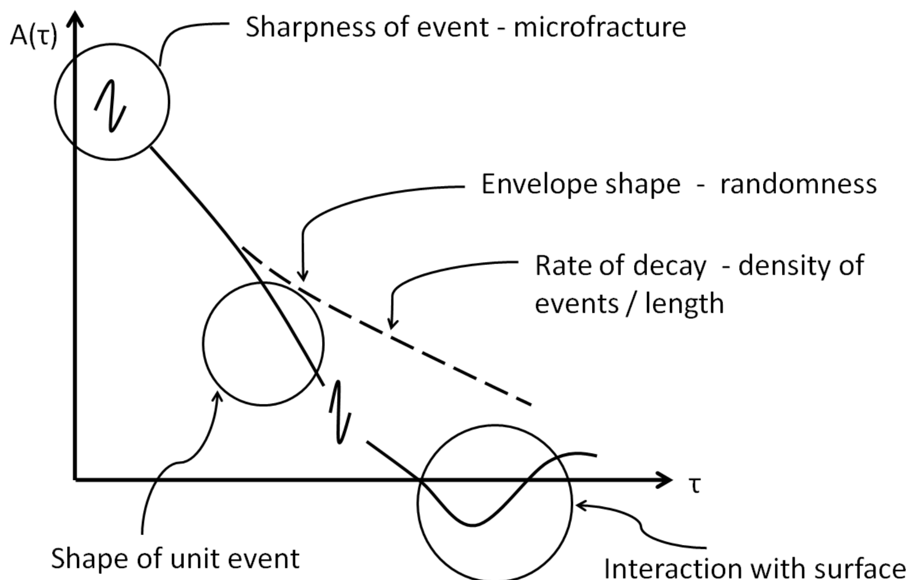


Figure 1.8 Relationship between ACF shape and grinding process [46].

1.3.2 Area ACF

The area ACF is also an important tool to diagnose stochastic surfaces. Compared to the linear ACF, the area ACF contains more information, for example, the surface texture direction.

1.3.2.1 Definition of Area ACF

The area autocovariance function (ACV) is defined as:

$$ACV(\tau_x,\tau_y)=\frac{1}{(m-\tau_x)(n-\tau_y)}\sum_{i=1}^{m-\tau_x}\sum_{j=1}^{n-\tau_y}z(i,j)z(i+\tau_x,j+\tau_y),\ \ 0\le\tau_x\le m-1,\ 0\le\tau_y\le n-1. \tag{1.17}$$

The area ACF is the area ACV normalized to the surface height mean square σ^2 :

$$ACF(\tau_x,\tau_y)=\frac{ACV(\tau_x,\tau_y)}{ACV(0,0)}=\frac{ACV(\tau_x,\tau_y)}{\sigma^2}. \tag{1.18}$$

1.3.2.2 Application of Area ACF

X. Jiang and D. J. Whitehouse have demonstrated the area ACFs for two milled surfaces under different machining conditions [5]. As Figure 1.9 shows, from the area ACF, it is possible to analyze the chatter in the milling process, which is one of the important factors that determine the surface quality.

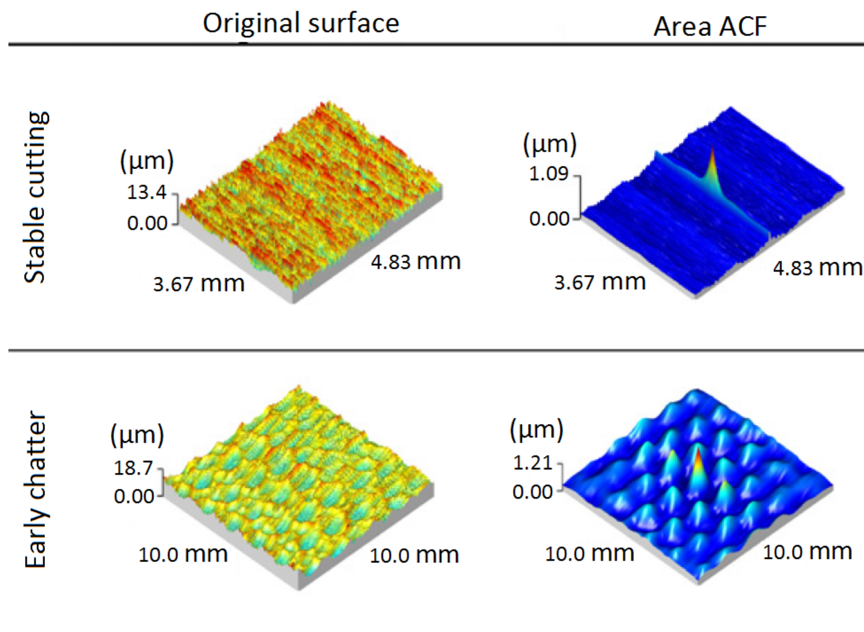


Figure 1.9 Comparison of the area ACFs for two milled surfaces (early chatter and stable cutting) [5].

1.4 Structure Function (SF)

Since the PSD and ACF have a number of disadvantages (PSD always needs additional mathematical processing such as windowing, zero padding and low order terms removal; ACF is strongly dependent on the mean plane and is not intuitive), the SF is a complementary method for analyzing the spatial frequency components.

1.4.1 Linear Structure Function of Both Profile and Area Data

The linear SF can be calculated from both profile data and areal data. It is the expectation of the squared height difference as a function of separation, which originates from the description of the astronomical seeing (atmospheric turbulence) by astronomers.

1.4.1.1 Astronomy

In astronomy, the image obtained by a telescope is affected by the atmospheric turbulence, the optics and the instrumentation. In order to achieve a high quality measurement, it is important to ensure that the telescope performance degradation arising from the optical errors is less than the best atmosphere from a statistical analysis [49]. Therefore, it is necessary to characterize the wavefront phase errors introduced by the atmospheric turbulence on various spatial scales; the SF is commonly chosen by the astronomers to do so.

The SF is calculated in terms of the averaged phase difference over all pairs of points in the phase map of a given separation. In the 1960's, it started with a model developed by Tatarski [50] and Fried [51] whose work was based on Kolmogorov [52] statistics. For Kolmogorov turbulence, the structure function, $S(r)$, can be expressed as

$$S(r) = \left\langle [z(r') - z(r'+r)]^2 \right\rangle = (\frac{\lambda}{2\pi})^2 6.88(\frac{r}{r_0})^{\frac{5}{3}}, \qquad (1.19)$$

where the angled brackets denote an average over r', $z(r')$ is the phase at position r', $z(r'+r)$ is the phase a distance r away from the point r', and r_0 is the correlation length that depends on the wavelength and scales to the 6/5 power (for a default wavelength of 0.5 μm) [49, 53]. The correlation length is also called the Fried parameter, indicating the strength of the atmospheric turbulence. The typical values are on the order of centimeters and larger values denote better images [49]. Figure 1.10 shows the SF for Kolmogorov turbulence with a correlation length of 0.1 m and a wavelength of 0.5 μm (given by R. E. Parks). If the spatial frequencies cover a wide range, it is usually plotted on a log/log scale.

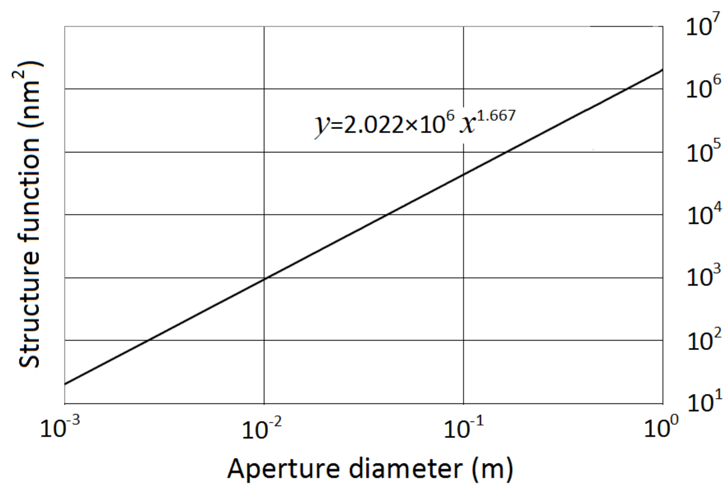


Figure 1.10 SF for Kolmogorov turbulence with a correlation length of 0.1 m and a wavelength of 0.5 μm (after [53]).

In the past few decades, the SF has been used as the polishing or figuring specification for large astronomical optics on various spatial scales. In the 1980's, it was first successfully applied as fabrication specifications for the William Herschel Telescope (WHT) [54,55]. Since then, the same specification strategy has been used for other large telescope mirrors, such as the Large Binocular Telescope (LBT) with two 8.4 meter

primary mirrors [56,57], the Giant Magellan Telescope [58,59,60,61,62], and the Thirty Meter Telescope (TMT) [63,64].

Figure 1.11 shows the linear SF analysis for a GMT segment (given by J. H. Burge, et al.). The mirror specification was derived from an atmospheric SF with r_0 = 92 cm. The allowable error after testing is given by the subtraction (in quadrature) of the optical test errors from the mirror specification. The result indicates that the optical test errors are smaller than the mirror specification, which means there is still room in the budget for errors on the mirror itself [49,58].

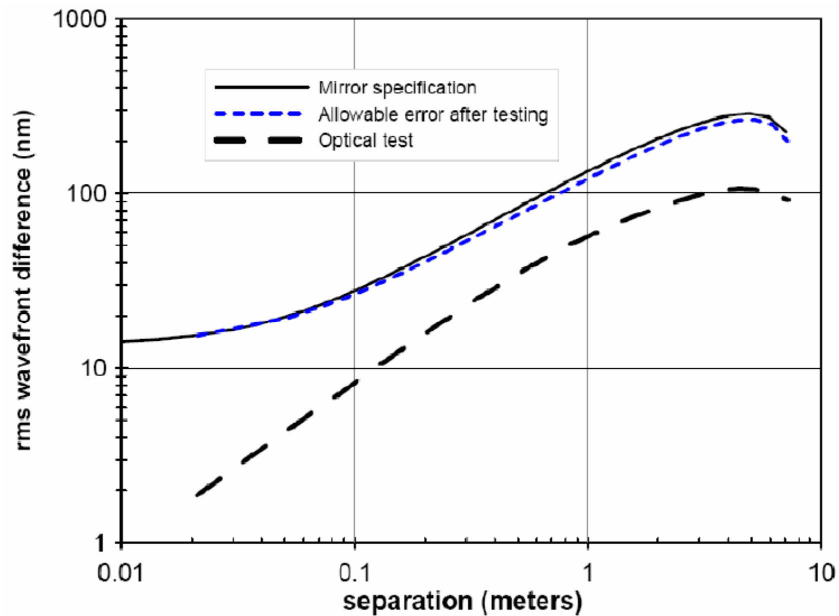


Figure 1.11 Linear SF analysis for a GMT segment [49,58].

1.4.1.2 Rough Surfaces

The SF was first introduced in the surface metrology (for profile data) in D. J. Whitehouse's PhD thesis [26] in 1971,

$$S(\tau) = E\left\{\left[z(x) - z(x+\tau)\right]^2\right\}, \tag{1.20}$$

where E is the statistical expectation and $z(x)$ is the value of the surface profile height z at x.

It is a rather brief appearance in a voluminous document written at a time when tribologists were focused on treating surfaces as random processes. D. J. Whitehouse pointed out that for a strictly stationary surface (rare in optics) the SF contains no more information than the ACF. Afterwards, many researchers (such as T. R. Thomas, C. A. Brown, and J. C. Russ) have applied the SF to surface finish characterization and fractal analysis [1,65,66,67,68,69]. For a fractal profile, the SF can be expressed as [65,66],

$$SF(\tau) = \Lambda^{2D-2}\tau^{2(2-D)}, \tag{1.21}$$

where τ is the shift, D is the fractal dimension (indicating how the roughness changes with sampling interval), and Λ is the topothesy which is the distance along the profile for which the expected angle between two points is one radian [66]. Usually the distance is far below the resolution of the measuring instruments and smaller than the atoms [66].

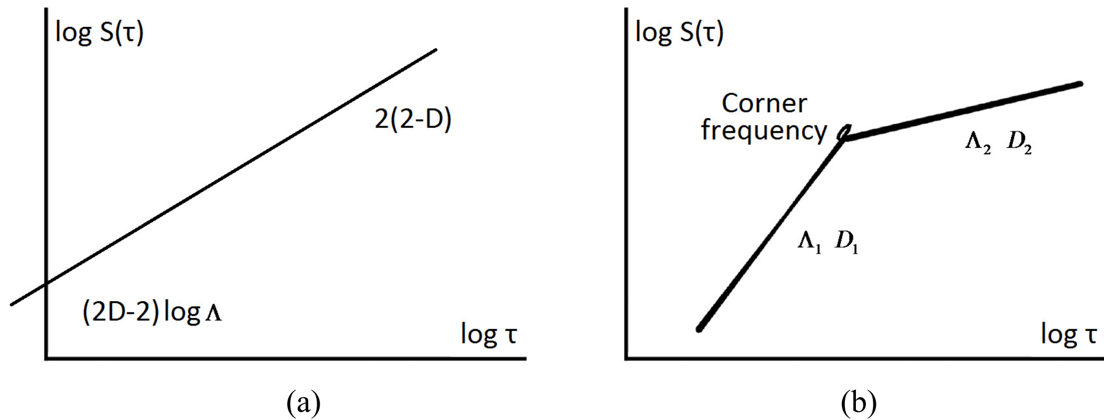


Figure 1.12 SF on log/log scale. (a) Estimation of fractal parameters using the SF [46]. (b) SF of a multifractal surface [71].

From Eq. (1.21), it is clear that the SF can be plotted as a straight line on log/log scale. As Figure 1.12 shows, the fractal dimension D and the topothesy Λ can be

obtained easily from the slope and intercept of this straight SF line. These parameters are intrinsic properties of the surface, which means they are independent of the measured length and the filtering techniques [70,71].

In practice, machined surfaces are often produced by several different processes each with its fractal properties in different spectral bands. Thus, they are called multifractal [66] surfaces, which usually depict a SF in the form of several straight lines with different slopes meeting at a discontinuity. The wavelength associated with the discontinuity can indicate the transition from one mechanism to another for the surface formation. As shown in Figure 1.12 (b), the transition point is called corner frequency [72,73].

T. R. Thomas, B.-G. Rosén and their coworkers have applied the SF to a cylinder liner. As Figure 1.13 shows, the data are based on the stylus and AFM measurements of worn and unworn cylinder linear surfaces with different sampling intervals, indicating multifractal characteristics and corner frequencies. When τ is approximately 20 μm, there is a transition which indicates the dimensions of the largest honing grits. The smaller features are generated by a single and continuous fractal process, corresponding to the fracture of the grits. Moreover, as a dimensionless number connecting roughness properties to the material property ratio, the plasticity index [74] can be obtained to identify different surfaces according to their wear resistance [71].

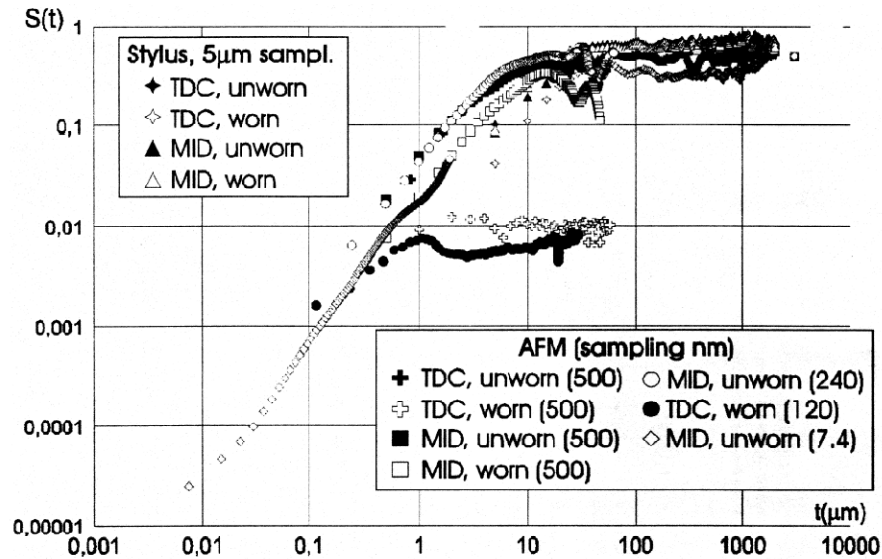


Figure 1.13 SF of a cylinder liner in worn and unworn conditions (MID=midpoint of piston stroke, TDC=top dead center) [75][76].

Besides the man-made surfaces, the SF has been applied to some naturally occuring surfaces. For example, C Y Poon has used the SF to analyze the surface of naturally fractured rocks [77]. Figure 1.14 depicts the SF against the delay length τ for limestone, sandstone, chert and carbonate. The results indicate that the SFs for all types of rock surfaces are consistent with Eq. (1.21) for the self-affine fractals. Thus, the characterization of the naturally fracture rocks requires two parameters - the fractal dimension D and the topothesy Λ , which can be calculated from the SF.

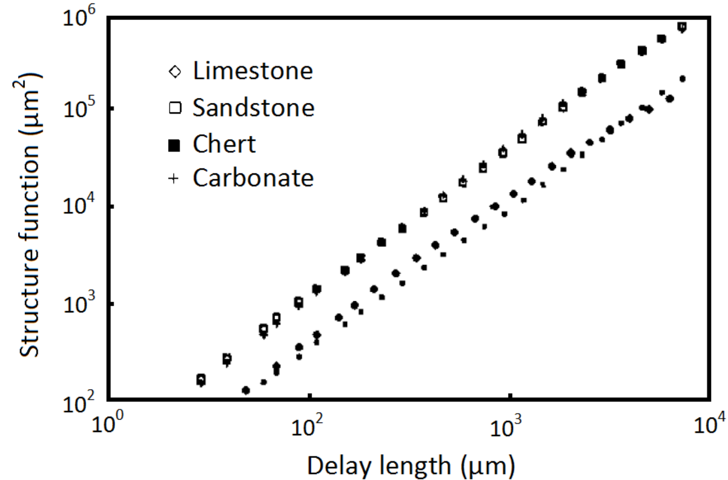


Figure 1.14 SF of surface profiles of limestone, sandstone, chert and carbonate [77].

1.4.1.3 General Optical Surfaces

Recently, R. E. Parks has applied the linear SF to a general optical surface [53]. Figure 1.15 (a) is the height map of an optical surface with a diameter of 100 mm, and Figure 1.15 (b) is the related linear SF calculated by R. E. Parks. The high value in the linear SF at 30 mm separation is due to the central high in the input map of Figure 1.15 (a), dropping to the low zone at a radius of about 30 mm. In other words, during the calculation, the average of uniformly sampled point pairs at separation of 30 mm includes a large number where one of the point pairs is in the central high and the other is in the annular low zone. Similarly, the high value in the linear SF at around 80 mm separation is due to the annular low zone and the high edge.

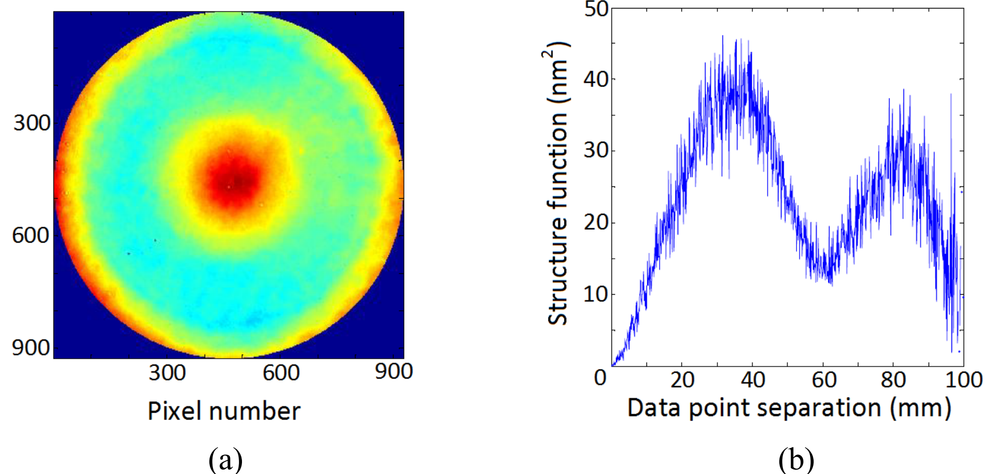


Figure 1.15 Linear SF of the general optical surface [53]. (a) Height map. (b) Linear SF calculated by R. E. Parks.

For better visualization over a wide spatial frequency range, the linear SF can be plotted on log/log scale in Figure 1.16. Note that the RMS of the linear SF (which can be interpreted as the RMS as a function of separation) is shown in green and the RMS slope, again as a function of separation, can easily be obtained from the SF [53]. Specifications can be plotted on the same axes in Figure 1.16. In this case, the optical part is better than the specification but not by a large margin.

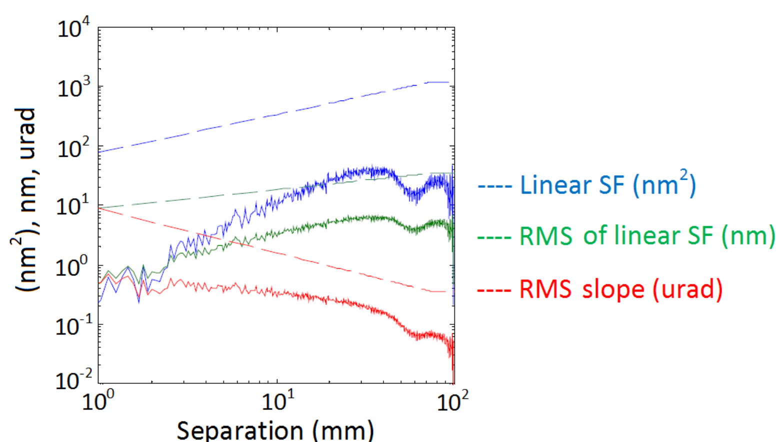


Figure 1.16 The linear SF, RMS and RMS slope of the wavefront in Figure 1.15 (a) and the respective specifications (after [53]).

1.4.2 Area Structure Function

The linear SF compresses area height information ($z(\tau_x,\tau_y)$ to 1D ($z(\tau)$), thus losing information on anisotropy. Figure 1.17 and Figure 1.18 show the calculated linear SFs and area SFs for two different surfaces generated with unit coefficients for two different Zernike terms. The two surfaces in Figure 1.17 (a) and Figure 1.18 (a) have markedly different azimuthal frequencies, but the two linear SFs are almost indistinguishable. In contrast, there is significant difference between the two area SFs. The area SF retains information on surface anisotropy.

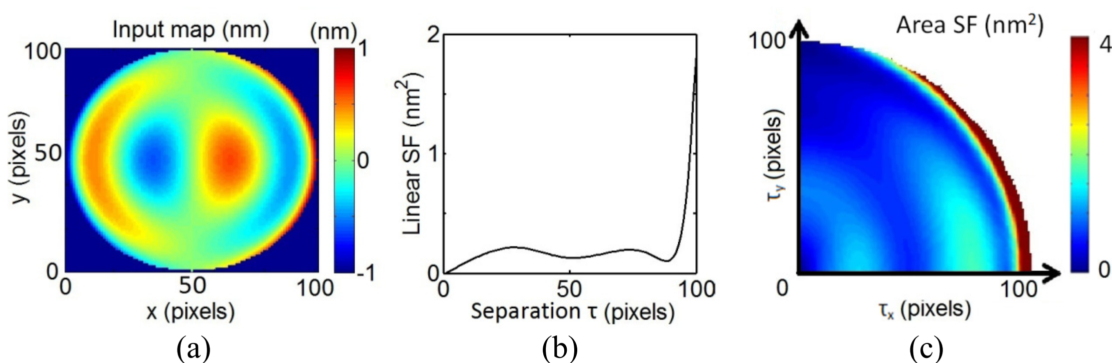


Figure 1.17 The linear SF and area SF of simulated map 1. (a) Simulated map 1. (b) Linear SF of map 1. (c) Area SF of map 1.

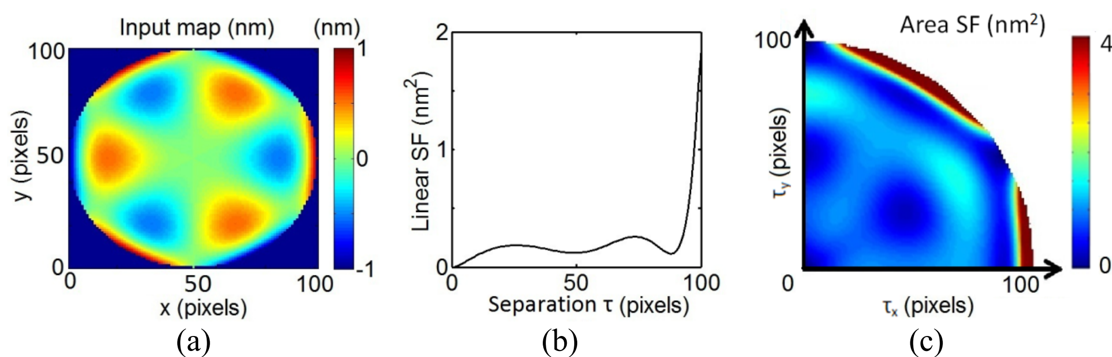


Figure 1.18 The linear SF and area SF of simulated map 2. (a) Simulated map 2. (b) Linear SF of map 2. (c) Area SF of map 2.

The area SF was first briefly introduced by R. S. Sayles and T. R. Thomas in 1977 in polar terms [25]. Two more recent publications show area SF in Cartesian coordinates for rough surfaces such as cylinder liners [70,71].

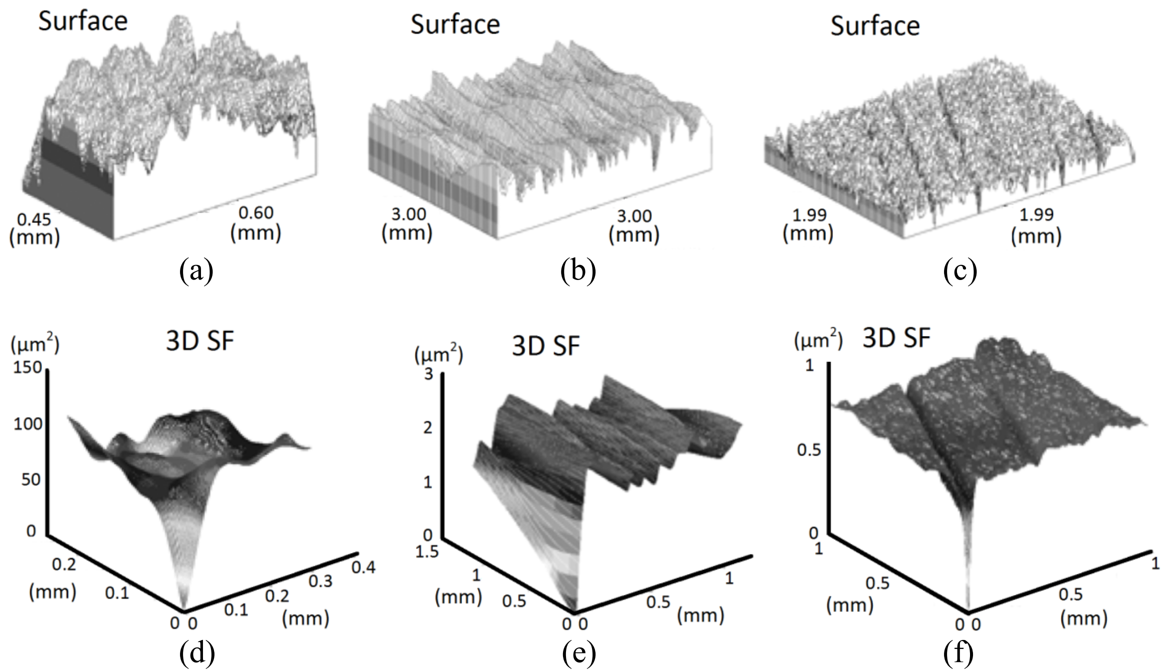


Figure 1.19 Three machined surfaces and their area SFs (after [70,71]). (a) Grit-blasted. (b) Surface-ground. (c) Plateau-honed. (d) Area SF of grit-blasted surface. (e) Area SF of ground surface. (f) Area SF of plateau-honed surface.

T. R. Thomas and his coworkers have demonstrated the area SFs for three different machined surfaces (Figure 1.19): grit-blasted, surface-ground and plateau-honed. As they pointed out, the fractal parameters are invariant with direction for an isotropic surface (e.g. grit-blasted surface). Highly anisotropic surfaces (e.g. ground surface) have a minimum in fractal dimension close to the lay, and the topothesy changes greatly in different directions of the surface [70,71].

1.5 Significance of This Project

From the definitions and applications of the PSD, ACF and SF, it is clear that the PSD and ACF have limitations (PSD always needs additional mathematical processing such as

windowing, zero padding and low order terms removal; ACF is strongly dependent on the mean plane and is not intuitive) in characterizing the spatial frequency components. The SF solves some of these problems and may be an alternative or complementary way of specifying and characterizing optical surfaces.

Although the linear SF has been used in surface metrology since the 1970s, most works are focused on the fractal characterization of surface roughness. We only found two papers [53,78] that discuss the application of linear SF to general optical surfaces. Thus, there are still enough opportunities to explore the linear SF for the optical surface characterization in spatial frequency domain.

However, the linear SF does not capture anisotropy on the surface. It is necessary to apply the area SF, which has not been widely used in surface metrology. A typical application is the fractal characterization for some machined surfaces (T. R. Thomas et al [70,71]).

Consequently, it is worthwhile to investigate both the linear and area SF for the optical surfaces characterization in the spatial frequency domain.

1.6 Layout of Dissertation

This dissertation investigates the structure function (SF) for optical surfaces characterization in spatial frequency domain.

Chapter 1 has introduced the motivation of this project. For optical surfaces and transmitted wavefront, it is important to characterize errors in the spatial frequency domain. Currently, the power spectral density function (PSD) and, less often, the auto-correlation function (ACF) have been used to describe the spatial frequency characteristics. The SF, especially the area SF offers some advantages.

Chapter 2 describes the calculation of the linear SF and the area SF. For large numbers of points within the aperture, computational time increases rapidly, so sampling is required. Comparisons, show that the sliding sampling we developed is faster and more accurate than the conventional sampling strategies. A two-quadrant area SF is introduced because the one-quadrant area SF does not completely describe surfaces with certain asymmetries. Next, the relationship between the linear SF and the two-quadrant area SF is investigated for circular apertures, showing that the linear SF is the average over concentric semi-circles of the two-quadrant area SF. Based on the analysis of periodic errors with both the linear SF and area SF, the spatial frequency content of a diamond turned surface is evaluated.

Chapter 3 investigates the relationship between the SF and other surface characterization techniques: Zernike polynomials, ACF, PSD, and RMS gradient.

It turns out that the linear SF of the sum of the Zernike terms equals to the sum of the linear SF of each Zernike term only under certain conditions.

For stationary surface, the SF contains similar information as ACF, but it provides better visualization of the surface characteristics than ACF, which is strongly dependent on the mean plane and is not intuitive.

The SF is computationally “correct” for any arbitrary aperture shape without extra processing, while the PSD always needs additional mathematical processing, such as windowing, zero padding, low-order terms removal, and choice of sub-apertures for irregular apertures.

After connecting the SF to the RMS gradient, the SF slope at the origin has been evaluated.

Chapter 4 demonstrates a methodology for area SF combination of data from instruments with substantially different lateral dynamic ranges. The area SF of the measured sub-aperture (representing the high spatial frequency content) depends on the measurement position in the full aperture, especially for anisotropic surfaces. To solve this problem, a sampling method of sub-aperture measurement has been developed. The correct ASF can be estimated by adding the correct tilt to each sampled sub-aperture, typically estimated from the large-area measurement with knowledge of the global coordinates of each sub-aperture measurement. Moreover, the effect of the instrument transfer function (ITF) of the instruments used has been analyzed.

Chapter 5 investigates other applications of the SF. For the phase shifting interferometry (PSI), the analysis with the SF shows that the air turbulence has more influence than the electronic noise of the interferometer, and the long interference cavity introduces more environmental effect than the short cavity. Particularly, an instantaneous interferometer was used to make dynamic measurement of the air turbulence. In addition, the instantaneous interferometer was applied to implement an on-machine measurement for a diamond turning machine, which provides an effective feedback for the compensation machining. Finally, the analysis with SF gives a quantitative estimation in the spatial frequency domain – the accuracy of the diamond turned surface has been improved after the compensation machining.

Chapter 6 presents the conclusions of this work and suggests some future work.

CHAPTER 2 : CALCULATION AND INTERPRETATION OF STRUCTURE FUNCTION

The structure function (SF) is the expectation of the squared height difference as a function of separation. A linear SF can be computed for profile or area data, as discussed in Chapter 1, with the loss of any detail of anisotropy in the surface. The area SF contains the anisotropic information of a surface and can be calculated without filtering, zero padding and restraints on aperture size or shape. Although the definition of the SF has been given, there are some different calculation techniques which lead to different computing speed. Especially for large numbers of points within the aperture, a variety of sampling methods may be used. Section 2.1 introduces the calculation of the linear SF and explores various sampling strategies. Section 2.2 develops a two-quadrant area SF because the one-quadrant area SF does not completely describe surfaces with certain asymmetries. Section 2.3 investigates the relationship between the linear SF and the area SF. Based on the analysis of periodic errors with both the linear SF and area SF, section 2.4 demonstrates the analysis of the spatial content of a diamond turned surface.

2.1 Linear Structure Function

Figure 2.1 shows the calculation of the linear SF for simulated area data. Conceptually all possible point pairs of separation τ_1 are selected, the squared height differences for each point pair calculated, and then averaged to obtain $SF(\tau_1)$. This process is repeated for all separations τ. The SF can be computed over any chosen dynamic range for any aperture shape. Separation into figure, mid-spatial frequency and roughness can still be

done, but it is not necessary. For large numbers of points within the aperture, computational time increases rapidly, so sampling is required.

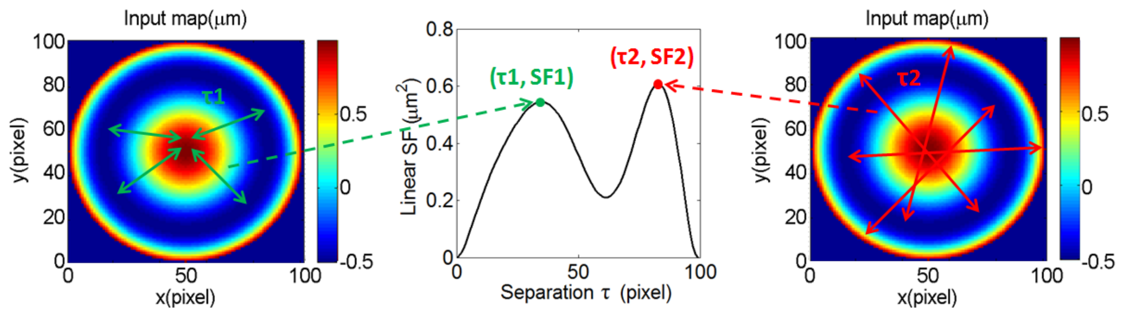


Figure 2.1 Calculation of linear SF of area data.

2.1.1 Paired Sampling

The paired sampling is easy to compute. As shown is Figure 2.2 (a), the first step is to randomly pick two points within the aperture, compute the distance between these two points as the separation τ of the structure function and calculate the squared height difference [53]. This process is repeated the number of times required to provide acceptable convergence of the computed SF, as seen for example in the noisiness of the plot or more rigorously in the variance for each separation. The average of the squared height differences for each separation τ gives the linear SF.

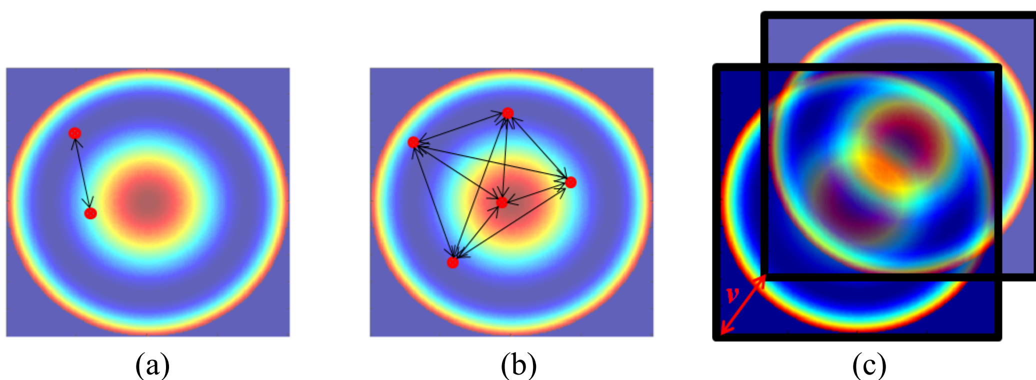


Figure 2.2 Sampling strategies. (a) Paired sampling. (b) Pure sampling. (c) Sliding sampling.

2.1.2 Pure Sampling

Here the process is to randomly pick N points (choosing N for adequate convergence) within the aperture, calculate the height differences over the all pairs of points, bin separations according to the required resolution, average the height differences along with the corresponding separation τ, and plot the linear SF. Figure 2.2 (b) shows an example with $N = 5$.

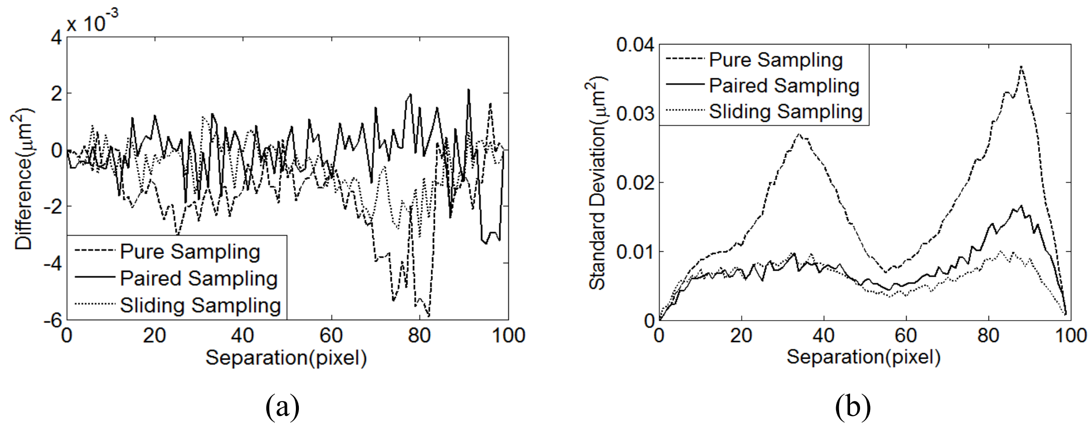
2.1.3 Sliding Sampling

The sliding sampling we developed is shown in Figure 2.2 (c). Here the process is to duplicate the height map, randomly offset it with the separation vector v, pick the points in the overlap region, calculate the squared height difference between the two maps at each point, bin the magnitudes of the vectors according to the required resolution as the separations, average the squared height differences with related separation, and obtain the linear SF.

2.1.4 Comparison

In order to investigate the accuracy and speed of the sampling strategies, these three methods were respectively applied to the same height data (Figure 2.1). For a single calculation of the SF, the computational time of each sampling method was set to be identical. Then the SFs were calculated 100 times for each sampling method. Figure 2.3 (a) shows the difference between the true value given by the full-data calculation and the mean value of 100 sampled SFs. Most of the differences of paired sampling and sliding sampling are limited to $\pm 2\times10^{-3}$ μm^2 (the range of the true SF values is from 0 to 0.607 μm^2), while the differences of pure sampling are greater. That is, paired sampling and

sliding sampling are more accurate than pure sampling. Moreover, as Figure 2.3 (b) shows, the sliding sampling has the lowest standard deviation of the 100 calculations.



(a) (b)

Figure 2.3 Comparison of sampling strategies. (a) Difference between the true value and the mean value of 100 sampled SFs. (b) Standard deviation of each sampling strategy.

In this case, sliding sampling has the highest accuracy at the same computational time. In other words, sliding sampling is the fastest method to achieve a given accuracy when compared to the other two sampling methods. The reason is that all the points in the overlap region are related to only one separation when sliding the duplicated map, so it is unnecessary to calculate each separation of the point pairs. Calculation using all points with sliding method is faster than paired and pure sampling when the number of sampled points is large. Under the same condition, the sliding calculation with all points can be faster than the sliding sampling calculation, because picking of random points takes increasingly more time as the sampling rate is increased. Consequently, there is a threshold point at which it is faster to calculate all points than to sample. However, for non-circular aperture, the accuracy of sliding method may decrease, because the same separation can be related to different overlapped regions (the number of data point in the overlapped regions are different). Thus, when averaging the squared height differences from different overlapped regions (with the same separation), the deviation may increase.

2.2 Two-quadrant Area Structure Function

The definition of area SF can be expressed as,

$$S(\tau_x,\tau_y)=\frac{1}{(m-\tau_x)(n-\tau_y)}\sum_{i=1}^{m-\tau_x}\sum_{j=1}^{n-\tau_y}\left\{z(i,j)-z(i+\tau_x,j+\tau_y)\right\}^2,\quad 0\le\tau_x\le m-1, 0\le\tau_y\le n-1,\tag{2.1}$$

where there are m equally spaced points in the x direction and n equally spaced points in the y direction; τ_x and τ_y are two integers; and $S(\tau_x,\tau_y)$ is the expectation of the squared height difference of points separated by τ_x and τ_y. Note that in Eq. (2.1) (given by Thomas et al. [71]) τ_x and τ_y are non-negative integers. Hence the computed SF represents only one "quadrant" in a Cartesian coordinate system (Figure 2.4). This is sufficient for surfaces which are rotationally invariant (or more generally are symmetric about either the x or y axis).

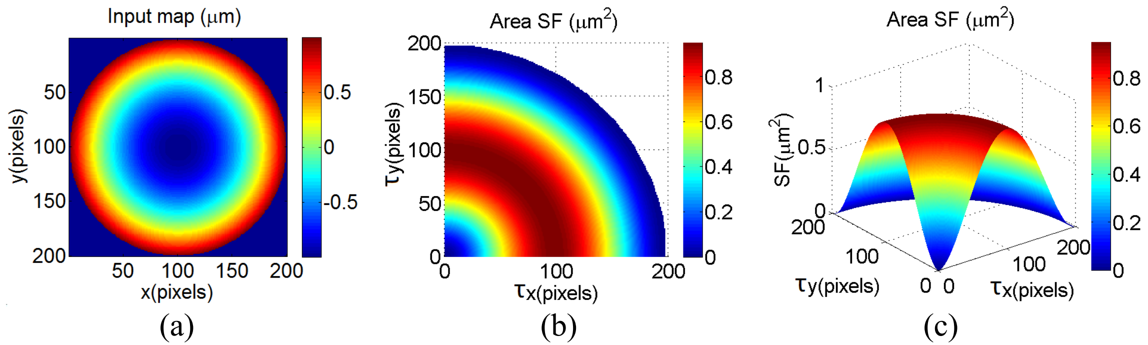


Figure 2.4 One-quadrant area SF. (a) Power surface. (b) One-quadrant area SF in 2-dimension. (c) One-quadrant area SF in 3-dimension.

However, for the asymmetric surfaces, the one quadrant representation of area SF is not sufficient. Thus, we developed the area SF in the second quadrant:

$$S(-\tau_x,\tau_y)=\frac{1}{(m-\tau_x)(n-\tau_y)}\sum_{i=1+\tau_x}^{m}\sum_{j=1}^{n-\tau_y}\left\{z(i,j)-z(i-\tau_x,j+\tau_y)\right\}^2,\quad 0\le\tau_x\le m-1, 0\le\tau_y\le n-1.\tag{2.2}$$

Figure 2.5 shows conceptually the calculation of the area SF according to Eqns. (2.1) and (2.2). The steps are: (1) duplicate the height map; (2) offset the duplicate map by (τ_x, τ_y) or $(-\tau_x, \tau_y)$, and overlay on the original map; (3) calculate the squared height differences in the overlap region; and (4) average to obtain the expectation, $S(\tau_x, \tau_y)$ or $S(-\tau_x, \tau_y)$.

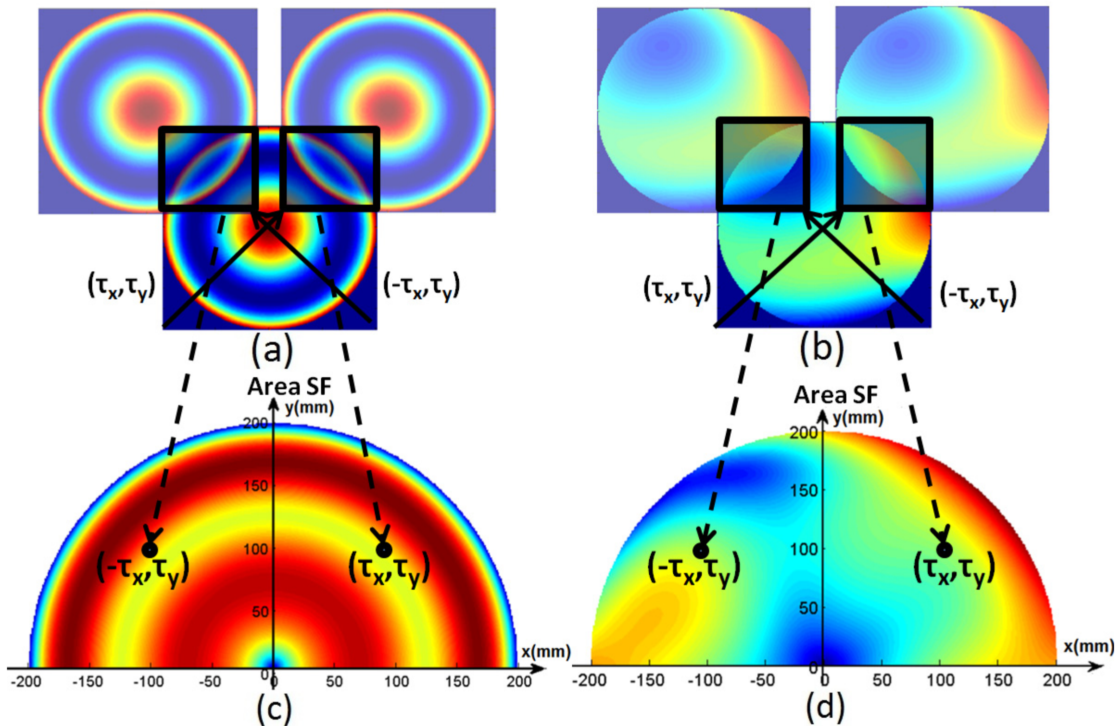


Figure 2.5 Calculation of the area SF. (a) Move the duplicated symmetric surface in 2 directions. (b) Move the duplicated asymmetric surface in 2 directions. (c) Area SF for the symmetric surface. (d) Area SF for the asymmetric surface.

Figure 2.5 (a), shows the calculation of the area SF in two directions (move the duplicated map in the upper-right and upper-left directions, respectively), and the result is shown in Figure 2.5 (c). Because of the symmetry of the surface, the area SF in the first quadrant contains the same information as that in the second quadrant ($S(\tau_x, \tau_y) = S(-\tau_x, \tau_y)$).

However, for asymmetric surfaces, Thomas' equation is insufficient, because in general $S(\tau_x, \tau_y)$ is not equal to $S(-\tau_x, \tau_y)$ (the second quadrant SF). For example, in Figure 2.5 (b) and (d), the area SF in the first quadrant is not the same as that in the second quadrant ($S(\tau_x, \tau_y) \neq S(-\tau_x, \tau_y)$). Because SF is the squared value, we can find $S(\tau_x, \tau_y) = S(-\tau_x, -\tau_y)$ and $S(-\tau_x, \tau_y) = S(\tau_x, -\tau_y)$. Therefore, for rotationally varying surfaces that are asymmetric about both *x* and *y* axes, the two-quadrant area SF is required for a complete description of the spatial characteristics of the surface.

By definition, the SF at the origin $(\tau_x = 0,\ \tau_y = 0)$ is always identically zero – that is the expectation of the squared height difference between the surface map and itself with no separation is zero. At small separations, the SF of a measured optical surface will represent three components; the measurement noise and the surface form and finish. Consider the case where the measurement noise is dominated by electronic noise (i.e. random noise at each pixel). The SF (squared expectation of the phase difference) of this noise is independent of separation. If the surface contains high spatial frequency components, this will be represented by all positive harmonics of that frequency (e.g. the SF of the sinusoidal surface). If the surface is dominated by low frequency components (Figure 2.5), the SF will be non-zero for small separations, increasing as the separation increases.

The SF analysis of a surface with figure error (e.g. low-order Zernike components) is useful for building physical understanding, as shown in Figure 2.5. Consider the two-quadrant SF graphs in Figure 2.5 (a) and (c). The input map to the SF calculation (Figure 2.5 (a)) is third order spherical aberration. At a separation equal to the radius, the SF tends toward zero since the majority of the calculable separations compare the central

peak with the edge. The annular high values in the area SF of Figure 2.5 (c) at 70 mm and 170 mm separation are due to the central high in the input map of Figure 2.5 (a), dropping to the low zone at a radius of about 70 mm and then the edge that is again high.

Figure 2.5 (b) and (d) represent a more realistic scenario in which the surface under test contains asymmetric components. In the first quadrant of Figure 2.5 (d), the SF value is high at the edge (large separation), indicating a significant height difference between the upper-right and lower-left edges of the surface. But for the same separation in the second quadrant, the SF values are relatively low on the edge, showing the similarity in the surface heights at the upper-left and lower-right. Clearly, the long spatial wavelength components of this surface have a strong directional component. The two quadrant area SF shows this directionality for all spatial wavelengths.

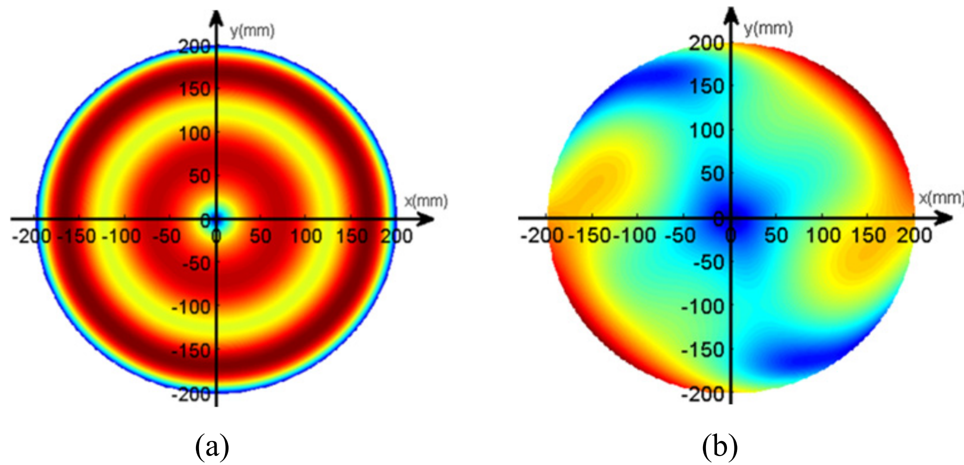
Also, the SF values are directly connected to height errors, a useful intuitive aspect for the optics manufacturer. The values are the surface variance as a function of separation, and the square root of the SF is the surface RMS as a function of separation.

As with all parametric descriptions of surface data, the area SF is biased in the presence of noise. The number of points averaged is a function of separation. For small τ, there is a large amount of averaging and the uncertainty from this source is relatively small. For larger separations, the uncertainty increases.

The area SF in two quadrants takes the shape of the half-aperture of regular optical apertures when τ_x and τ_y are plotted on linear axes. The two quadrants are symmetric for some but not all surface symmetries.

For better physical interpretation in some cases, a four-quadrant representation may be used (Figure 2.6). Quadrant three contains exactly the same information as quadrant one,

and it can be computed by flipping quadrant one about *x*- and *y*- axes. The same applies to quadrants two and four.



(a) (b)

Figure 2.6 The area SFs in four quadrants. (a) Four-quadrant area SF of the symmetric surface in Figure 2.5 (a). (b) Four-quadrant area SF of the asymmetric surface in Figure 2.5 (b).

For large numbers of points within the aperture, computational time increases rapidly, so sampling is required. Similar to the linear SF, the sliding sampling method is used for the area SF: Duplicate the height map, randomly offset it with the separation vector (τ_x, τ_y) or $(-\tau_x, \tau_y)$, pick the points in the overlap region, calculate the squared height difference between the two maps at each point, bin the vectors according to the required resolution as the separations, average the squared height differences with related separation vectors, and obtain the area SF.

2.3 Relationship between Linear Structure Function and Area Structure Function

For circular aperture, the linear SF can always be calculated from the area SF calculated for both quadrants, although the inverse is clearly not true. Conceptually, $\tau = \sqrt{\tau_x^2 + \tau_y^2}$ for all values of *x* and *y*. Expressed differently, the linear SF is the average over concentric semi-circles of the two-quadrant area SF.

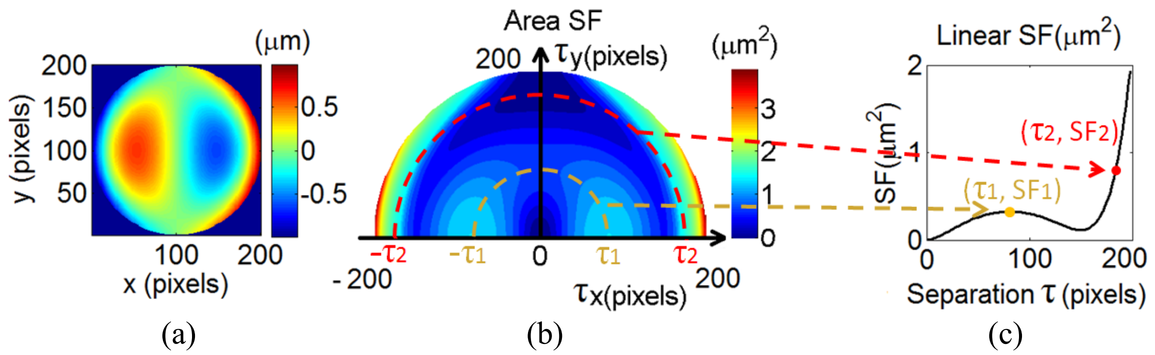


Figure 2.7 Relationship between the linear SF and the area SF. (a) Input height map. (b) Area SF of the map. (c) Linear SF of the map.

For example, Figure 2.7 shows the area SF and Linear SF for a surface (coma). The point (τ_1, SF_1) in the linear SF curve is related to the average SF over a concentric semi-circle with a radius of τ_1 in the area SF. So is the point (τ_2, SF_2).

Particularly, if the input map is rotationally invariant, the related area SF is also rotationally invariant. That is, the SF values are constant for values of τ_x and τ_y for which $\tau = \sqrt{\tau_x^2 + \tau_y^2}$ is constant. Because the linear SF value is based on the corresponding separation in every possible direction, any profile in this area SF from the center to the edge is identical to the linear SF. For example, a radial profile extracted from the area SF in Figure 2.8 (b) is equal to the linear SF in Figure 2.8 (c).

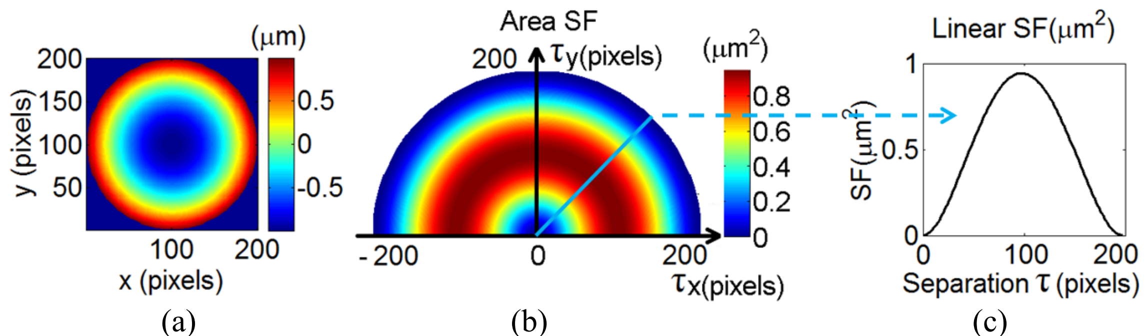


Figure 2.8 Relationship between the linear SF and the area SF. (a) Input height map. (b) Area SF of the map. (c) Linear SF of the map.

2.4 Spatial Frequency Interpretation with Structure Function

In modern precision surface fabrication, a variety of processes can lead to periodic errors. For example, in single point diamond turning or small tool polishing, slideway error motions and misalignment, tool wear, and vibrations can generate periodic errors over a broad range of frequencies. The SF provides an alternative way of representing spatial frequency content compared to the PSD and ACF. The SF of a periodic surface is also periodic, although not exactly so for sampled data. Maxima and minima occur in the SF at harmonics of the periodic height information. This effect can easily be seen in SF calculations of simulated height data that is sinusoidal.

2.4.1 Simulated Periodic Errors

2.4.1.1 Sinusoidal Profile

Figure 2.9 (a) shows 1D simulated height data that is a sinusoidal function with the spatial wavelength of 10000 pixels, and Figure 2.9 (b) shows the calculated linear SF.

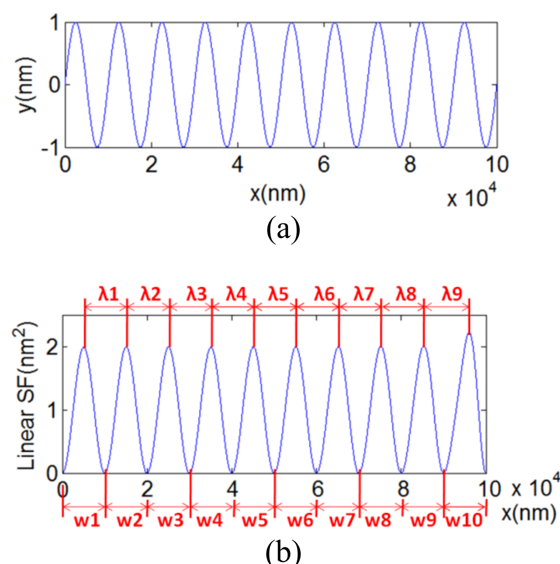


Figure 2.9 Linear SF for 1D sinusoidal height data. (a) Simulated 1D sinusoidal height data. (b) Linear SF of the simulated data.

Because the SF is the average of the squared height differences, the periodic maximum SF values are close to the separation values with the maximum height differences (Table 2.1). The maxima are not exactly separated by the spatial wavelength because of finite sampling and averaging effects. The periodic minimum SF values are, however, exactly separated by the spatial wavelength, because SF values are identically zero when the separation is exactly an integral multiple of the wavelength.

Table 2.1 Parameters of the linear SF in Figure 2.9.

	1	2	3	4	5	6	7	8	9	10
λ (nm)	10007	10008	10010	10014	10021	10032	10057	10135	10662	
w (nm)	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000
SF(nm^2)	2.0006	2.0007	2.0009	2.0012	2.0017	2.0025	2.0041	2.0081	2.0228	2.2321

If the spatial wavelength is calculated from the periodic minimum SF values, it is clear that the SF oscillates with the same spatial frequency as the raw data. This can be useful for extracting the average periodic length scale on a surface.

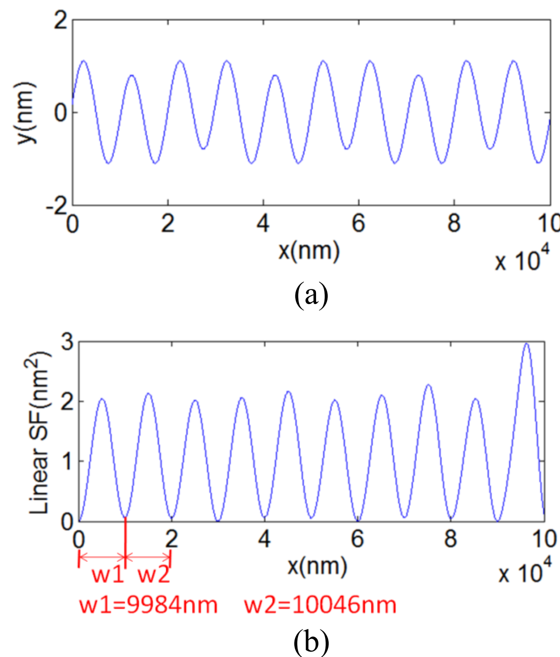
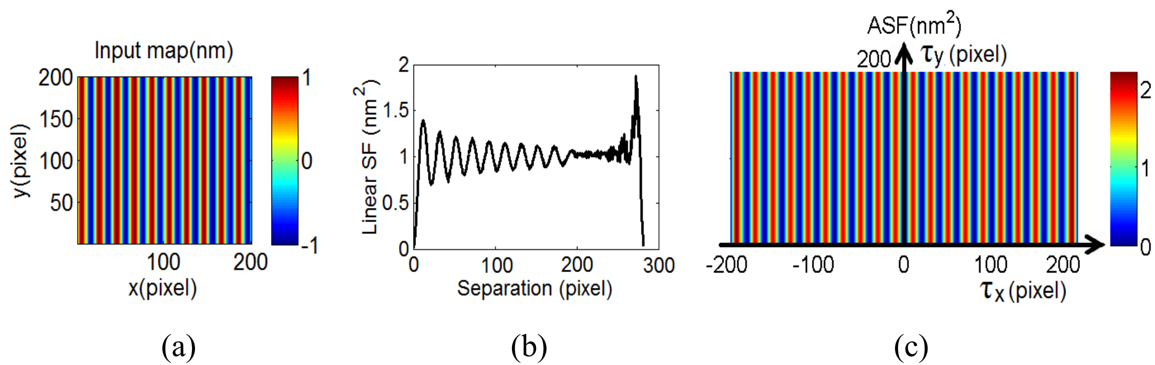


Figure 2.10 Linear SF for 1D combined sinusoidal height data. (a) 1D combined sinusoidal height data. (b) Linear SF of the height data.

Since the SF is not an orthogonal representation of the spatial content of a surface, some computational issues should be taken into consideration. For example, in Figure 2.10, after combining another low-frequency sinusoidal profile (wavelength = 30000 pixels, amplitude = 0.2 nm, initial phase = 2/3 π) to the sinusoidal profile in Figure 2.9 (a), the SF shows that even the wavelengths calculated from the periodic minimum SF values are not exactly the same. In such cases, the average high-frequency periodic length scale is best calculated by first removing the low frequency component (e.g. form) before calculating the SF.

2.4.1.2 Sinusoidal Surface

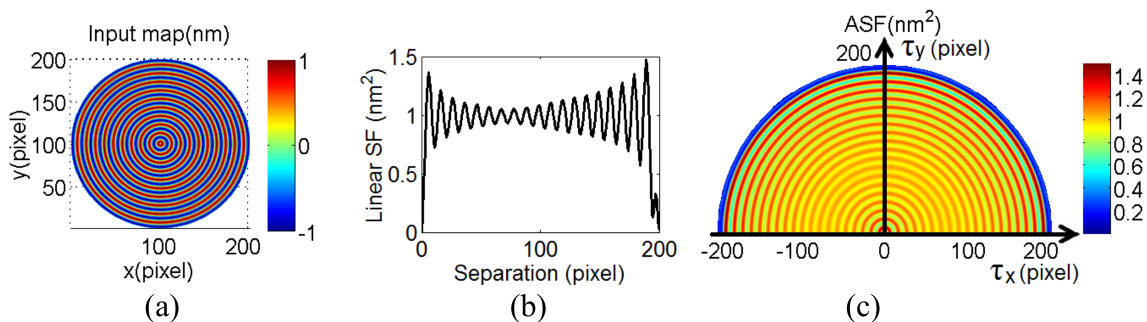


(a) (b) (c)

Figure 2.11 SFs for 2D linear sinusoidal height data. (a) Simulated 2D linear sinusoidal height data. (b) Linear SF of the simulated data. (c) Area SF of the simulated data.

Figure 2.11 shows the SFs of a linear sinusoidal surface. Note that the linear SF in Figure 2.11 (b) contains significant noise when the separation is greater than 200 pixels. This is because the number of points available to calculate the average SF value at these large separations dramatically decreases. Unlike the profile data, the linear SF maxima values, based on the areal data, gradually decrease because of the averaging in different directions of the surface. In contrast, the area SF in Figure 2.11 (c) is not impacted by this

effect because the directional information is not averaged – the area SF includes spatial frequency data in all directions.



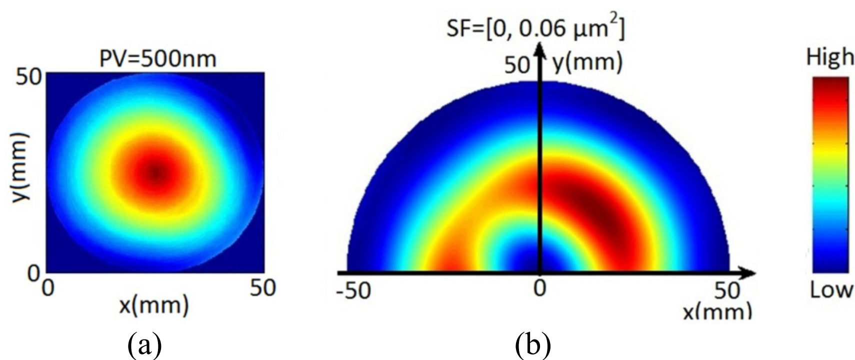
(a) (b) (c)

Figure 2.12 SFs for 2D radial sinusoidal height data. (a) Simulated 2D radial sinusoidal height data. (b) Linear SF of the simulated data. (c) Area SF of the simulated data.

Figure 2.12 shows the SFs of a simulated radial sinusoidal surface. Similar to Figure 2.11 (b), the maxima SF values also change with the separation values. With a round aperture, the linear SF is only calculated for separations less than 200 pixels, so the noise impact of few points at very large separations seen in Figure 2.11 (b) is not present.

The ripple of the area SF in Figure 2.12 (c) is rotationally invariant, as discussed in Figure 2.8, any profile in this area SF from the center to the edge is identical to the linear SF in Figure 2.12 (b).

2.4.2 Diamond Turned Aluminum



(a) (b)

Figure 2.13 Analysis for diamond turned aluminum surface. (a) Measured surface height map. (b) Area SF of the surface height map.

Figure 2.13 (a) shows the analysis for a Fizeau interferometer measurement of a diamond-turned aluminum flat. The area SF of the input map (Figure 2.13 (b)) is dominated by 2 features: an orientation-independent peak at a separation of approximately 25 mm, and an asymmetry due to astigmatism in the part. The 25 mm peak is due to the conical shape to the part, a common form error when there is a squareness error in the machine. The astigmatism is probably not caused by a machine error, and more likely the result of mount-induced deformations.

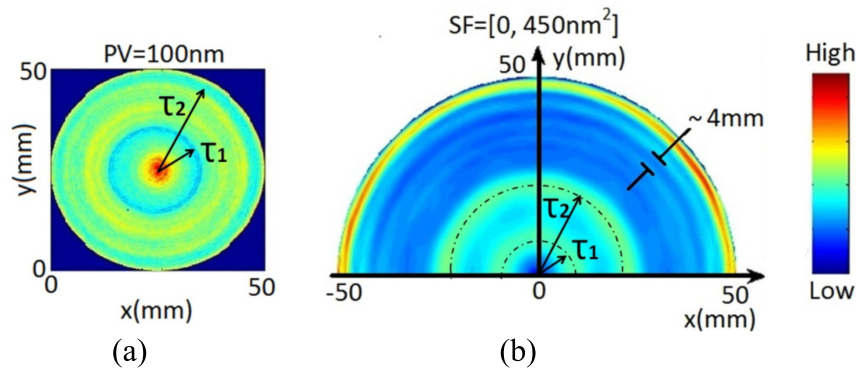


Figure 2.14 Analysis for diamond turned aluminum surface. (a) Residual error after removing the first 36 Zernike terms. (b) Area SF of the residual error.

If mid-spatial frequency error is the primary concern, the SF analysis is best carried out after removing the form error. Here 36 Zernike terms are removed to obtain the residual error, as shown in Figure 2.14 (a). The area SF calculation for the residual error is shown in Figure 2.14 (b). The annuli in the area SF at 10 mm and 22 mm separation are primarily due to the central high in the residual error (Zernikes do not fit a cone well), dropping to the low zone at a radius of about 10 mm, and then the edge that is low again.

Note the presence of a high frequency “ripple” in the area SF with a separation repeat length of about 4 mm. Like the sinusoidal example, this indicates a dominant mid-spatial frequency component in the surface at approximately 4 mm. Furthermore, the “ripple” is

most clearly seen in the region of the area SF plot corresponding to large separation. This can be appreciated from the linear SF analysis shown in Figure 2.12 (b), where the SF peak values are large for large separations. On the other hand, the SF peak values are also large for small separations in Figure 2.12 (b), but these SF peaks are not clear in the area SF in Figure 2.14 (b). This is because there are also two high spatial frequency periodicities on the surface and these lead to strong annuli peaks in the area SF at small separations. This shadows the high-frequency "ripple" that is also in the area SF data at these separations.

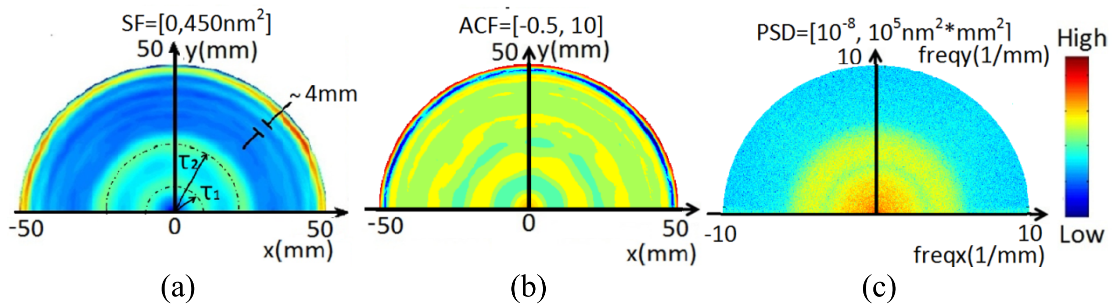


Figure 2.15 Analysis for diamond turned aluminum surface. (a) Area SF of the residual error map. (b) Area ACF of the residual error map. (c) Area PSD of the residual error map.

Figure 2.15 shows the analysis of the diamond-turned aluminum flat after removing the 36 Zernike terms. Comparison of Figure 2.15 (a) and (b) suggests that the area SF provides a clear visualization of the spatial content of the surface as a function of separation and orientation compared to the ACF, perhaps a result of a product vs. a difference as a function of separation. As Sayles and Thomas [25] point out, because the area ACF is the sum of terms each of which is the product of two amplitudes, it is less intuitive. Figure 2.15 (c) shows a PSD analysis of the data. The values between different spatial frequencies are not as clear as the area SF. Moreover, the PSD analysis is sensitive to computational details such as filtering, windowing and zero padding.

2.5 Summary

This chapter describes the calculation of the linear SF and the area SF. For large numbers of points within the aperture, computational time increases rapidly, so sampling is required. Comparisons show that the sliding sampling we developed is faster and more accurate than the conventional sampling strategies. A two-quadrant area SF is introduced because the one-quadrant area SF does not completely describe surfaces with certain asymmetries. Afterwards, the relationship between the linear SF and the two-quadrant area SF is investigated for circular apertures, showing that the linear SF is the average over concentric semi-circles of the two-quadrant area SF. Based on the analysis of periodic errors with both the linear SF and area SF, the spatial content of a diamond turned surface is evaluated.

CHAPTER 3 : CHARACTERISTICS OF STRUCTURE FUNCTION AND RELATIONSHIP TO OTHER PARAMETERS

As noted in Chapter 1, the spatial content of surface topography over broad spatial scales has been characterized by techniques derived from signal processing – for example PSDs and ACFs. In addition, specific descriptors such as the Abbott-Firestone (bearing area) curve and scale sensitive (tiling) analyses have been developed. This chapter explores in more detail the relationship among SFs, PSDs and ACFs. In the specific case of optical surfaces, low order aberrations can be related to appropriate combinations of Zernike polynomials, while the scattering from the surface is related to slope. These surface descriptions are also considered in terms of the SF. Section 3.1 defines the Zernike polynomials used in this dissertation, shows the SF of the individual Zernike polynomial terms, and investigates the SF of a surface described by multiple Zernike polynomials. It turns out that the linear SF of the sum of the Zernikes equals to the sum of the linear SF of each of Zernikes with different azimuthal frequencies. This theorem does not apply to area SFs. Section 3.2 describes the relationship between the SF and the ACF. Section 3.3 compares the SF to the PSD. Section 3.4 connects the SF to the RMS gradient and evaluates the SF slope at the origin.

3.1 Structure Function and Zernike Polynomials

Zernike polynomials [79,80,81] are commonly used to characterize and visualize circular apertures on optical surfaces [82], and are being used for the specification and tolerancing for some optics. Note, for example, that the Thirty Meter Telescope (TMT)

[63,64] optical specification uses a SF, but allows for removal of specified amounts of power, coma and astigmatism, etc (defined as combinations of Zernikes). Therefore, it is useful to explore the total SF for a surface described by Zernike polynomials. Burge et al have made some investigation on this topic [49,83], and their analysis is updated here and extended to the ASF. Other polynomial sets (see ISO 14999) may be used for other aperture shapes (rectangular, annuli, etc) [84], but such applications are rarer and will not be considered further here.

3.1.1 Zernike Polynomials

The Zernike polynomials are a sequence of polynomials that are orthogonal over the interior of the unit disk[85], which can be expressed as

$$Z_n^m(\rho,\theta) = R_n^m(\rho)\cos(m\theta), \qquad (3.1)$$

$$Z_n^{-m}(\rho,\theta) = R_n^m(\rho)\sin(m\theta), \qquad (3.2)$$

for the sine-cosine pairs of non-rotationally symmetric terms [82] and, for the rotationally symmetric terms

$$Z_n^0(\rho,\theta) = R_n^0(\rho), \qquad (3.3)$$

where n and m are nonnegative integers ($n - m \geq 0$ and even). The index n is the radial degree or the order of the polynomial because it indicates the highest power of ρ in the polynomial [86], and m is related the azimuthal frequency. ρ is the radial distance ($0 \leq \rho \leq 1$), and θ is the azimuthal angle between 0 and 2π radians. The radial polynomials $R_n^m(\rho)$ are defined as

$$R_n^m(\rho) = \sum_{k=0}^{(n-m)/2} \frac{(-1)^k (n-k)!}{k!(\frac{n+m}{2}-k)!(\frac{n-m}{2}-k)!}\rho^{n-2k}. \qquad (3.4)$$

Table 3.1 Zernike polynomials in Polar coordinates (the Fringe set (Loomis) [87])

#	Sets	Orthogonal polynomials	Aberration names
0	Z(0,0)	1	Piston
1	Z(1,1)	$\rho \cos\theta$	Tilt x
2	Z(1,-1)	$\rho \sin\theta$	Tilt y
3	Z(2,0)	$2\rho^2 - 1$	Power
4	Z(2,2)	$\rho^2 \cos 2\theta$	Astigmatism at 0°
5	Z(2,-2)	$\rho^2 \sin 2\theta$	Astigmatism at 45°
6	Z(3,1)	$(3\rho^2 - 2)\rho \cos\theta$	Coma x
7	Z(3,-1)	$(3\rho^2 - 2)\rho \sin \theta$	Coma y
8	Z(4,0)	$6\rho^4 - 6\rho^2 + 1$	Spherical
9	Z(3,3)	$\rho^3 \cos 3\theta$	Trefoil
10	Z(3,-3)	$\rho^3 \sin 3\theta$	Trefoil
11	Z(4,2)	$(4\rho^2 - 3)\rho^2 \cos 2\theta$	
12	Z(4,-2)	$(4\rho^2 - 3)\rho^2 \sin 2\theta$	
13	Z(5,1)	$(10\rho^4 - 12\rho^2 + 3)\rho \cos\theta$	
14	Z(5,-1)	$(10\rho^4 - 12\rho^2 + 3)\rho \sin \theta$	
15	Z(6,0)	$20\rho^6 - 30\rho^4 + 12\rho^2 - 1$	
16	Z(4,4)	$\rho^4 \cos 4\theta$	
17	Z(4,-4)	$\rho^4 \sin 4\theta$	
18	Z(5,3)	$(5\rho^2 - 4)\rho^3 \cos 3\theta$	
19	Z(5,-3)	$(5\rho^2 - 4)\rho^3 \sin 3\theta$	
20	Z(6,2)	$(15\rho^4 - 20\rho^2 + 6)\rho^2 \cos 2\theta$	
21	Z(6,-2)	$(15\rho^4 - 20\rho^2 + 6)\rho^2 \sin 2\theta$	
22	Z(7,1)	$(35\rho^6 - 60\rho^4 + 30\rho^2 - 4)\rho \cos \theta$	
23	Z(7,-1)	$(35\rho^6 - 60\rho^4 + 30\rho^2 - 4)\rho \sin \theta$	
24	Z(8,0)	$70\rho^8 - 140\rho^6 + 90\rho^4 - 20\rho^2 + 1$	
25	Z(5,5)	$\rho^5 \cos 5\theta$	
26	Z(5,-5)	$\rho^5 \sin 5\theta$	
27	Z(6,4)	$(6\rho^2 - 5)\rho^4 \cos 4\theta$	
28	Z(6,-4)	$(6\rho^2 - 5)\rho^4 \sin 4\theta$	
29	Z(7,3)	$(21\rho^4 - 30\rho^2 + 10)\rho^3 \cos 3\theta$	
30	Z(7,-3)	$(21\rho^4 - 30\rho^2 + 10)\rho^3 \sin 3\theta$	
31	Z(8,2)	$(56\rho^6 - 105\rho^4 + 60\rho^2 - 10)\rho^2 \cos 2\theta$	
32	Z(8,-2)	$(56\rho^6 - 105\rho^4 + 60\rho^2 - 10)\rho^2 \sin 2\theta$	
33	Z(9,1)	$(126\rho^8 - 280\rho^6 + 210\rho^4 - 60\rho^2 + 5)\rho \cos\theta$	
34	Z(9,-1)	$(126\rho^8 - 280\rho^6 + 210\rho^4 - 60\rho^2 + 5)\rho \sin \theta$	
35	Z(10,0)	$252\rho^{10} - 630\rho^6 + 560\rho^6 - 210\rho^4 + 30\rho^2 - 1$	
36	Z(12,0)	$924 \rho^{12} - 2772r^{10} + 3150r^8 - 1680r^6 + 420r^4 - 42r^2 + 1$	

There are various ordering rules of the Zernike polynomials [88,89,90], while the most commonly used set is that adopted by Loomis [87] (as shown in Table 3.1). In Table 3.1, the term $Z_n^{\pm m}$ is described in the form of *Z(n,±m)*.

If the Zernike polynomials are defined as normalized, the normalization term is given as

$$N_n^m = \sqrt{(2-\delta_{m0})(n+1)}, \tag{3.5}$$

where

$$\delta_{m0} = \begin{cases} 0, & \text{if } m \neq 0 \\ 1, & \text{if } m = 0 \end{cases}. \tag{3.6}$$

Because the Zernikes have an important property - orthogonality, the normalized Zernikes can also be called as orthonormal Zernikes.

In the radial part, the orthogonality reads

$$\int_0^1 \sqrt{(2-\delta_{m0})(n+1)} R_n^m(\rho) \sqrt{(2-\delta_{m'0})(n'+1)} R_{n'}^m(\rho) \rho d\rho = \delta_{nn'}, \tag{3.7}$$

where $\delta_{nn'}$ is the Kronecker delta

$$\delta_{nn'} = \begin{cases} 0, & \text{if } n \neq n' \\ 1, & \text{if } n = n' \end{cases}. \tag{3.8}$$

In the angular part ($m \neq 0$), the orthogonality can be represented as follows

$$\int_0^{2\pi} \cos(m\theta)\cos(m'\theta) d\theta = \pi(1+\delta_{m0})\delta_{mm'}, \tag{3.9}$$

$$\int_0^{2\pi} \sin(m\theta)\sin(m'\theta) d\theta = \pi\delta_{mm'}, \tag{3.10}$$

$$\int_0^{2\pi} \cos(m\theta)\sin(m'\theta) d\theta = 0, \tag{3.11}$$

$$\int_0^{2\pi} \sin(m\theta)\cos(m'\theta) d\theta = 0. \tag{3.12}$$

When it is integrated over the unit disk with both indices ρ and θ , the product of the radial and angular parts represents the orthogonality of the Zernike polynomials:

$$\sqrt{(2-\delta_{m0})(n+1)}\sqrt{(2-\delta_{m'0})(n'+1)}\int_0^1\int_0^{2\pi} Z_n^m(\rho,\theta)Z_{n'}^{m'}(\rho,\theta)\rho d\rho d\theta = \pi\delta_{jj'}, \quad (3.13)$$

where j and j' are the ordering numbers of the Zernike polynomials.

The optical surface, or the aberrations, may be described as the summation of orthonormal Zernikes

$$W(\rho,\theta) = \sum_{n=1}^{\infty}\sum_{m=1}^{n} c_n^m N_n^m Z_n^m(\rho,\theta) + \sum_{n=1}^{\infty}\sum_{m=1}^{n} c_n^{-m} N_n^{-m} Z_n^{-m}(\rho,\theta) + \sum_{n=0}^{\infty} c_n^0 N_n^0 Z_n^0(\rho,\theta), \quad (3.14)$$

where c_n^m is the coefficient associated with a particular term.

Because of the orthogonality and Eq. (3.14), the coefficients of the normalized Zernike polynomials fitted to the form represent the RMS of that term. Thus, the RMS of a surface constructed from a sum of normalized Zernike polynomials can be obtained when the coefficients are added in quadrature. Moreover, if some Zernike terms are subtracted or added from the surface, the values of other coefficients are not affected.

In addition to the definition in Polar coordinates $Z_n^m(\rho,\theta)$, the Zernike polynomials can be described in Cartesian coordinates ($Z_n^m(x,y)$), where $x=\rho\cos\theta$, $y=\rho\sin\theta$, $\rho^2 = x^2 + y^2$, and $0\le\rho\le 1$.

3.1.2 Linear SF of Zernike Polynomials

Since Zernike polynomials are widely used in optical system, it is useful to investigate the SFs for surface errors in terms of Zernike polynomials. Figure 3.1shows the Zernike polynomials as defined in Eqns. (3.1), (3.2) and (3.3) without normalization. And their linear SFs are discussed in the following section.

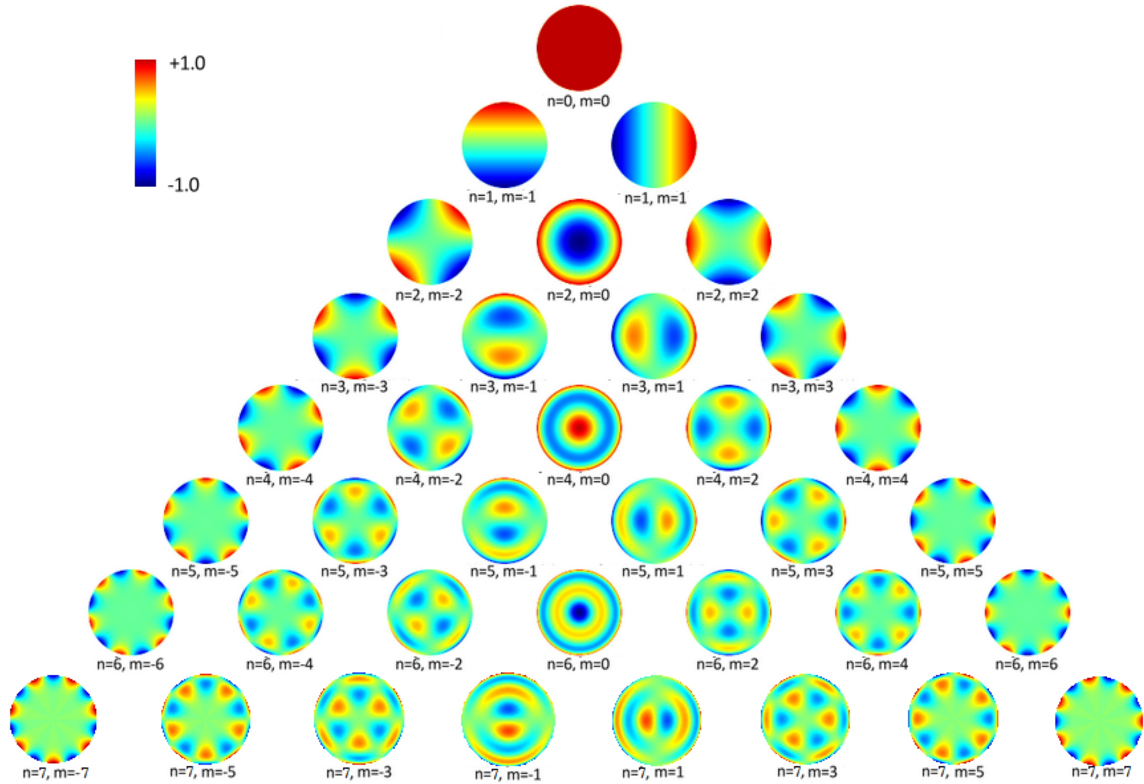


Figure 3.1 Zernike polynomials ($n \leqslant 7$, $m \leqslant 7$).

3.1.2.1 Linear SF of Individual Zernike Polynomial

Figure 3.2 describes the linear SFs of the Zernike polynomials in Figure 3.1. Because the linear SF is based on all possible pairs of points in the aperture, it is not affected by the rotation of the surface. As Figure 3.1 shows, all the rotationally variant Zernikes terms come in pairs only with a difference of rotation, thus both of them can be described by one linear SF since the linear SF averages azimuthally.

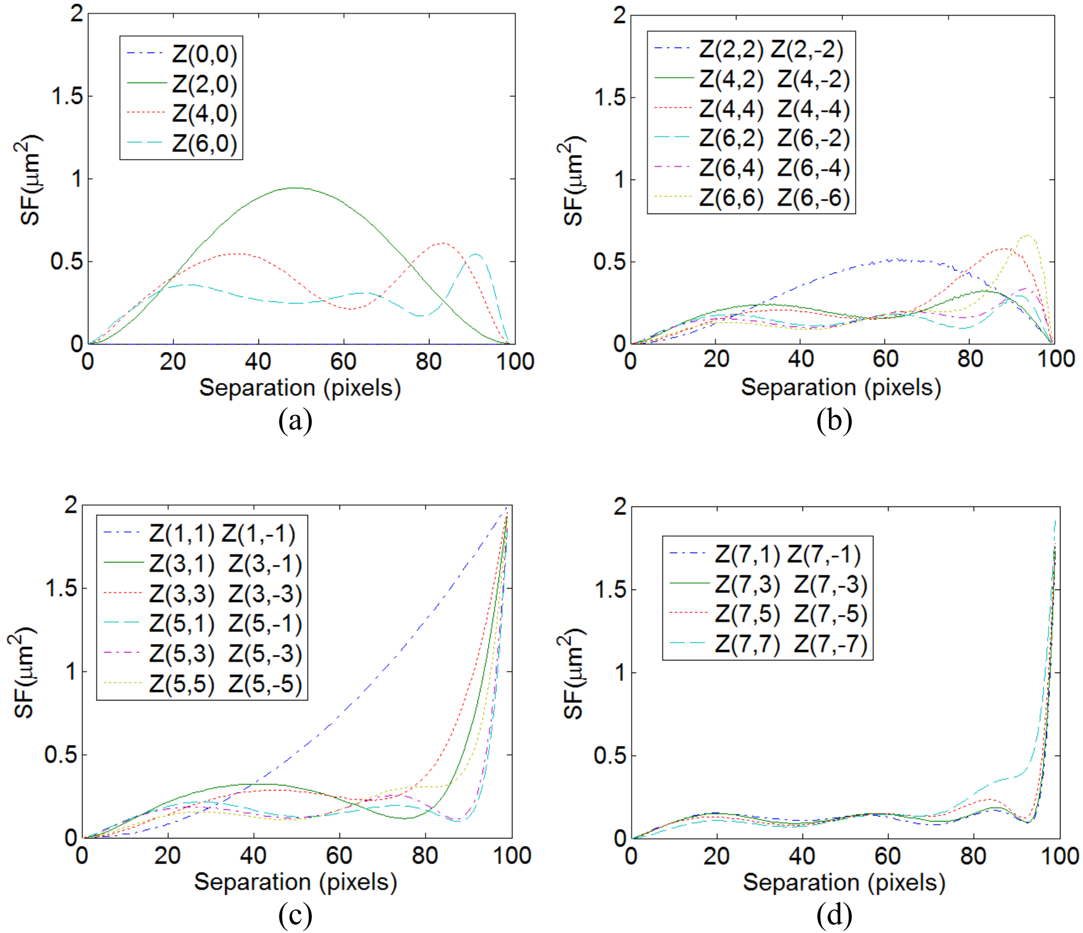


Figure 3.2 Linear SFs of the Zernikes. (a) Linear SFs of the rotationally invariant terms. (b) Linear SFs of the rotationally variant terms with even *m* values. (c) and (d) Linear SFs of the rotationally variant terms with odd *m* values.

With unit coefficients of the Zernikes (i. e. without normalization), all the even terms ((n,0), (n,2), (n,4) …) have a zero linear SF at maximum separation ($\rho = 1$):

$$SF = \frac{1}{2\pi}\int_0^{2\pi}\left[\cos(m\theta) - \cos(m(\theta+\pi))\right]^2 d\theta = \frac{1}{2\pi}\int_0^{2\pi} 0\, d\theta = 0, \quad (m = 0, 2, 4, 6...) \qquad (3.15)$$

or

$$SF = \frac{1}{2\pi}\int_0^{2\pi}\left[\sin(m\theta) - \sin(m(\theta+\pi))\right]^2 d\theta = \frac{1}{2\pi}\int_0^{2\pi} 0\, d\theta = 0. \quad (m = 0, 2, 4, 6...) \qquad (3.16)$$

With unit coefficients of the Zernikes, all the odd terms ((n,1), (n,3), (n,5) …) have the same linear SF (SF=2) at maximum separation ($\rho = 1$):

$$SF = \frac{1}{2\pi}\int_0^{2\pi}\left[\cos(m\theta) - \cos(m(\theta+\pi))\right]^2 d\theta = \frac{1}{2\pi}\int_0^{2\pi}\left[2\cos(m\theta)\right]^2 d\theta = 2, \quad (m = 1,3,5,7\ldots) \tag{3.17}$$

or

$$SF = \frac{1}{2\pi}\int_0^{2\pi}\left[\sin(m\theta) - \sin(m(\theta+\pi))\right]^2 d\theta = \frac{1}{2\pi}\int_0^{2\pi}\left[2\sin(m\theta)\right]^2 d\theta = 2. \quad (m = 1,3,5,7\ldots) \tag{3.18}$$

However, in Figure 3.2, the linear SFs are not exactly equal to 2 at maximum separation. That is because the calculation here is based on the discrete data, and only a few points are used at the edge (maximum separation). If the pixels density of the Zernikes matrix are increased, the linear SF at maximum separation will approach 2.

3.1.2.2 Total Linear SF of a Surface Described by Zernike Polynomials

The surface errors are usually described by a set of Zernike polynomials, therefore, it is necessary to explore the total linear SF of the surface errors.

Assume a surface is described by two different Zernike polynomials

$$Z(x,y) = Z_1(x,y) + Z_2(x,y), \tag{3.19}$$

which can also be expressed as

$$Z(P) = Z_1(P) + Z_2(P). \qquad P = (x,y) \tag{3.20}$$

Therefore, the linear SF is given as

$$S(\tau) = E\left\{\left[Z(P+\tau) - Z(P)\right]^2\right\} = E\left\{\{[Z_1(P+\tau) + Z_2(P+\tau)] - [Z_1(P) + Z_2(P)]\}^2\right\}, \tag{3.21}$$

where $Z(P+\tau)$ is the surface height at a distance τ from the position P.

$$S(\tau) = E\left\{\{[Z_1(P+\tau) - Z_1(P)] + [Z_2(P+\tau) - Z_2(P)]\}^2\right\} \tag{3.22}$$

$$\begin{aligned} S(\tau) = {} & E\left\{[Z_1(P+\tau) - Z_1(P)]^2\right\} + E\left\{[Z_2(P+\tau) - Z_2(P)]^2\right\} \\ & + 2E\left\{[Z_1(P+\tau) - Z_1(P)][Z_2(P+\tau) - Z_2(P)]\right\} \end{aligned} \tag{3.23}$$

$$S(\tau) = S_1(\tau) + S_2(\tau) \\ + 2E\{Z_1(P+\tau)Z_2(P+\tau)\} - 2E\{Z_1(P+\tau)Z_2(P)\} - 2E\{Z_1(P)Z_2(P+\tau)\} + 2E\{Z_1(P)Z_2(P)\} \qquad (3.24)$$

Because of the orthogonality of the Zernike polynomials in Eq. (3.13), that is, the integral of the product of two different Zernike polynomials is equal to zero:

$$E\{Z_1(P)Z_2(P)\} = E\{Z_1(P+\tau)Z_2(P+\tau)\} = 0. \qquad (3.25)$$

For $E\{Z_1(P)Z_2(P+\tau)\}$, the radial part is

$$\frac{1}{A}\int_0^1 \sqrt{2(n_1+1)}R_{n_1}^{m_1}(\rho)\sqrt{2(n_2+1)}R_{n_2}^{m_2}(\rho+l)\rho d\rho \neq 0, \qquad (3.26)$$

The angular part can be

$$\int_0^{2\pi} \cos(m_1\theta)\cos(m_2(\theta+\varphi))d\theta = \int_0^{2\pi} \cos(m_1\theta)[\cos(m_2\theta)\cos(m_2\varphi) + \sin(m_2\theta)\sin(m_2\varphi)]d\theta \\ = \cos(m_2\varphi)\int_0^{2\pi} \cos(m_1\theta)\cos(m_2\theta)d\theta + \sin(m_2\varphi)\int_0^{2\pi} \cos(m_1\theta)\sin(m_2\theta)d\theta = 0, \quad (if\ m_1 \neq m_2) \qquad (3.27)$$

or

$$\int_0^{2\pi} \sin(m_1\theta)\sin(m_2(\theta+\varphi))d\theta = 0, \quad (if\ m_1 \neq m_2) \qquad (3.28)$$

or

$$\int_0^{2\pi} \cos(m_1\theta)\sin(m_2(\theta+\varphi))d\theta = 0, \quad (if\ m_1 \neq m_2) \qquad (3.29)$$

or

$$\int_0^{2\pi} \sin(m_1\theta)\cos(m_2(\theta+\varphi))d\theta = 0, \quad (if\ m_1 \neq m_2) \qquad (3.30)$$

The product of the radial part and angular part is equal to zero when $m_1 \neq m_2$. In other words, the term $E\{Z_1(P)Z_2(P+\tau)\}$ is equal to zero when $m_1 \neq m_2$. And the term $E\{Z_1(P+\tau)Z_2(P)\}$ is in the same situation.

Therefore, the total linear SF is the sum of the individual linear SFs of each of the Zernike polynomials with different azimuthal frequencies:

$$S(\tau) = S_1(\tau) + S_2(\tau). \qquad (3.31)$$

Figure 3.3, Figure 3.4, and Figure 3.5 give a demonstration of the total lineal SF of two Zernike polynomials with different azimuthal frequencies. The result is consistent with the Eq. (3.31).

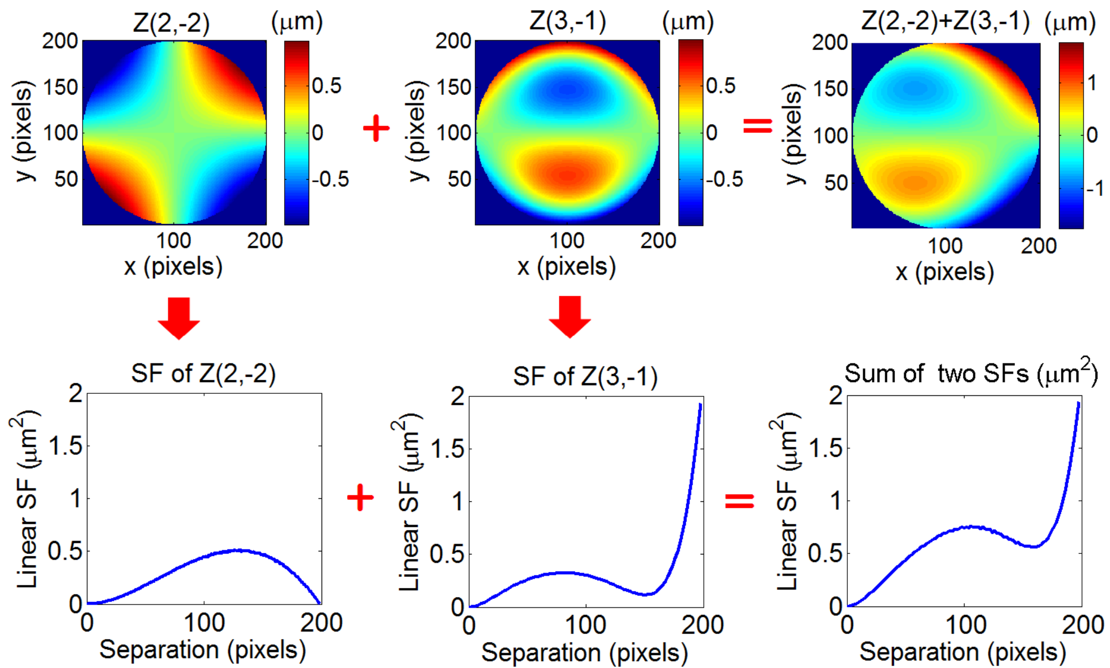


Figure 3.3 Sum of the two linear SFs of each of the Zernikes with different azimuthal frequencies.

As Figure 3.3 shows, *Z*(2, -2) and *Z*(3, -1) have different azimuthal frequencies. After calculating the linear SFs of *Z*(2, -2) and *Z*(3, -1) respectively, the sum of the two SFs is obtained.

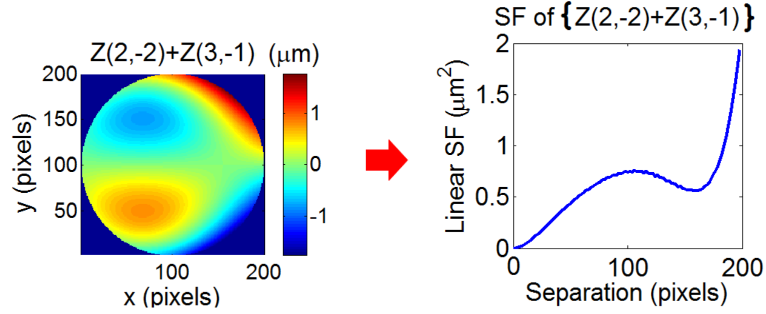


Figure 3.4 Linear SF of the sum of Zernikes.

Figure 3.4 shows the linear SF of the total Zernikes $\{Z(2, -2) + Z(3, -1)\}$. When differencing the SF of the total surface to the sum of the two SFs, the result is almost equal to zero (Figure 3.5). In theory, the difference is zero. In this case, because the calculation is based on the discrete data, the difference is not exactly equal to zero, but the values are so small that they can be ignored.

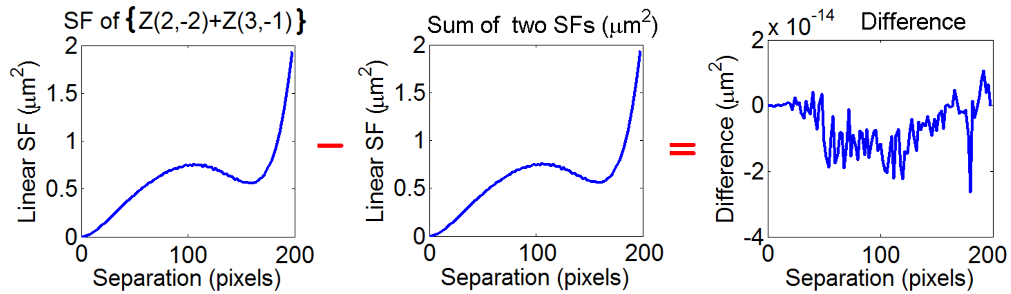


Figure 3.5 Difference between the SF of the total Zernikes and the sum of the two SFs.

If the Zernike polynomials are scaled by some factor, for example, $Z_{total}(P) = cZ(P)$, then the SF can be described as

$$S_{total}(\tau) = E\left\{\left[cZ(P+\tau) - cZ(P)\right]^2\right\} = c^2 S(\tau). \qquad (3.32)$$

In conclusion, if the surface is described by the Zernikes (with different azimuthal frequencies) in the form of

$$Z(P) = c_1 Z_1(P) + c_2 Z_2(P) + c_3 Z_3(P) + ..., \qquad (3.33)$$

then the total linear SF can be expressed as

$$S(\tau) = c_1^2 S_1(\tau) + c_2^2 S_2(\tau) + c_3^2 S_3(\tau)... \,. \qquad (3.34)$$

For example, a surface is given as

$$Z(P) = 2\,Z(2,-2) + 3\,Z(3,-1), \qquad (3.35)$$

where the two individual Zernikes have been shown in Figure 3.3, and the combined surface is shown in Figure 3.6 (a). Figure 3.6 (b) is the linear SF of the combined surface, and Figure 3.6 (c) is the sum of the individual SFs which have been scaled by the square of the related coefficients. The difference between Figure 3.6 (b) and (c) is approximately zero, the small values are due to the calculation which is based on the discrete data. Thus, the result is consistent with Eq. (3.34).

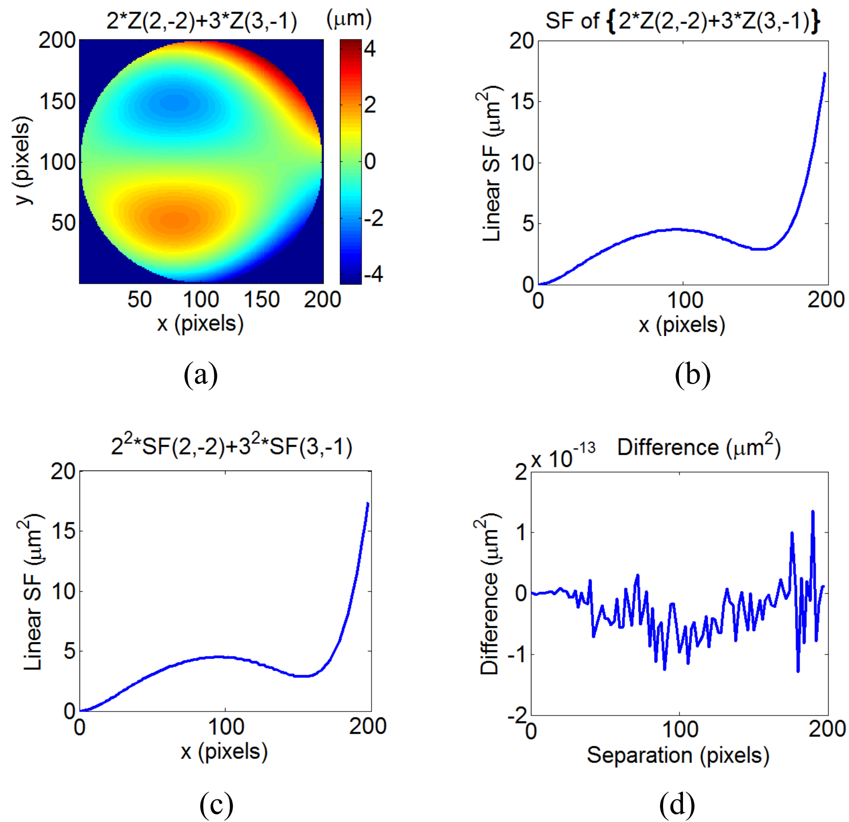


Figure 3.6 Combination of the linear SFs with coefficients. (a) Combined surface. (b) Linear SF of the combined surface. (c) Sum of the individual SFs with scaled coefficients. (d) Difference between (b) and (c).

However, if the combined Zernikes have the same azimuthal frequency, Eqns. (3.31) and (3.34) do not apply.

Figure 3.7, Figure 3.8, and Figure 3.9 give a demonstration of the total lineal SF of two Zernike polynomials with the same azimuthal frequency. The result shows that the sum of each linear SF is not equal to the linear SF of the sum of the Zernike polynomials.

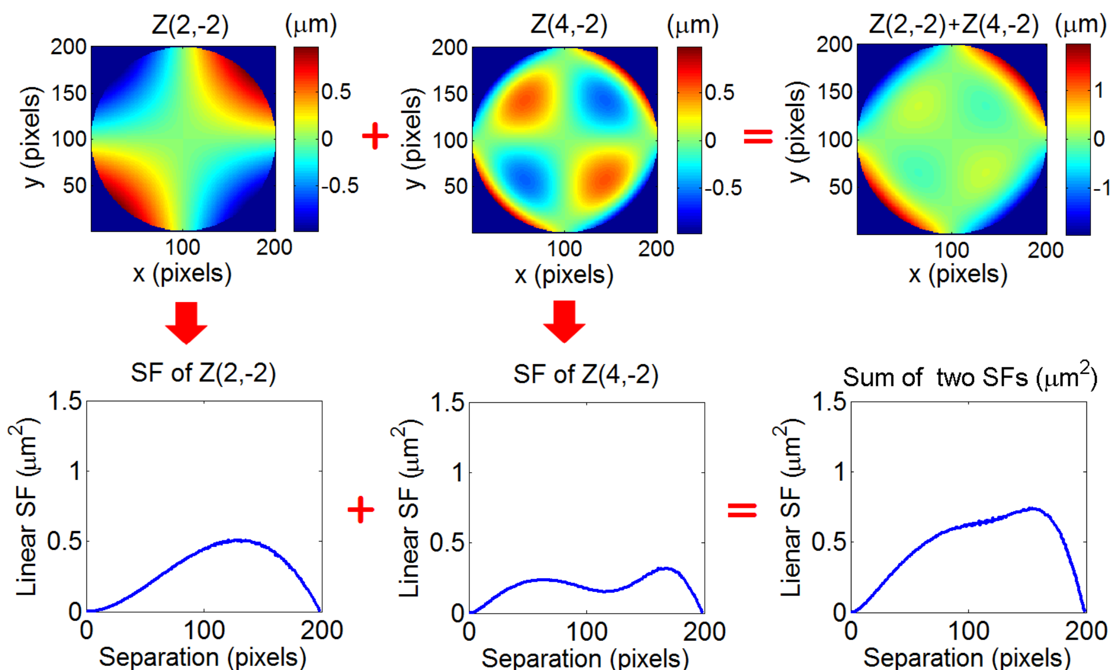


Figure 3.7 Sum of the two linear SFs of each of the Zernikes with the same azimuthal frequency.

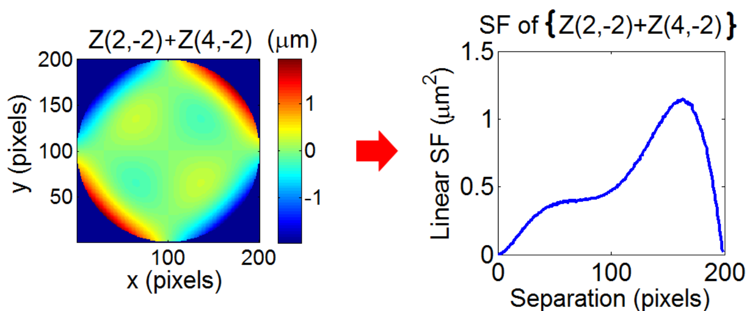


Figure 3.8 Linear SF of the total Zernikes.

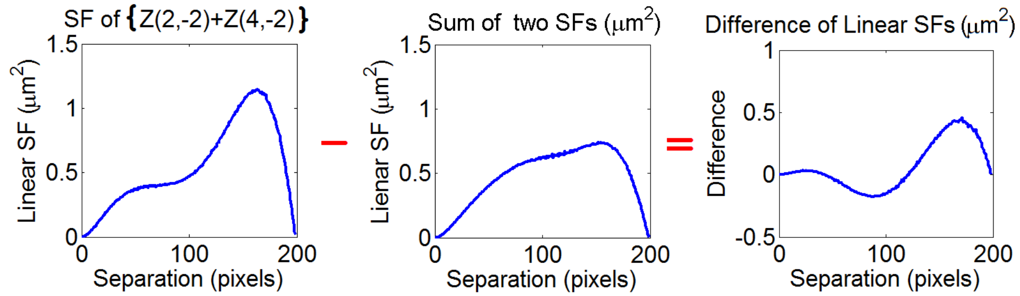


Figure 3.9 Difference between the SF of the total Zernikes and the sum of the two SFs.

3.1.3 Area SF of Zernike Polynomials

Since the linear SF does not capture anisotropy on the surface, it is necessary to investigate the area SF which can be calculated for any aperture shape, for all spatial content and anisotropies.

3.1.3.1 Area SF of Individual Zernike Polynomial

Figure 3.10 to Figure 3.14 describe the area SFs for astigmatism, coma, trefoil, defocus and third order spherical respectively.

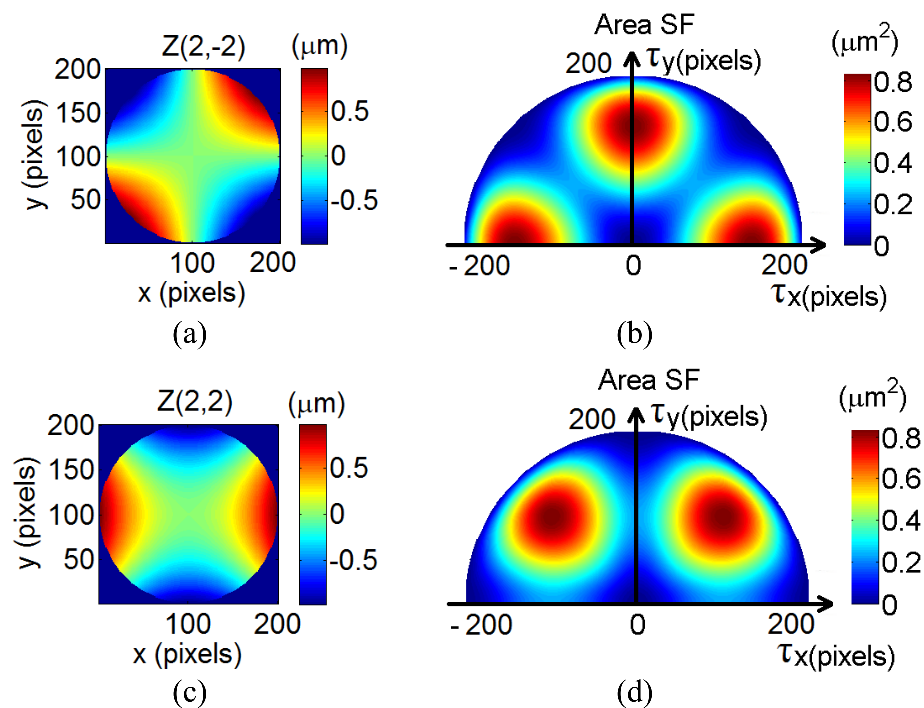
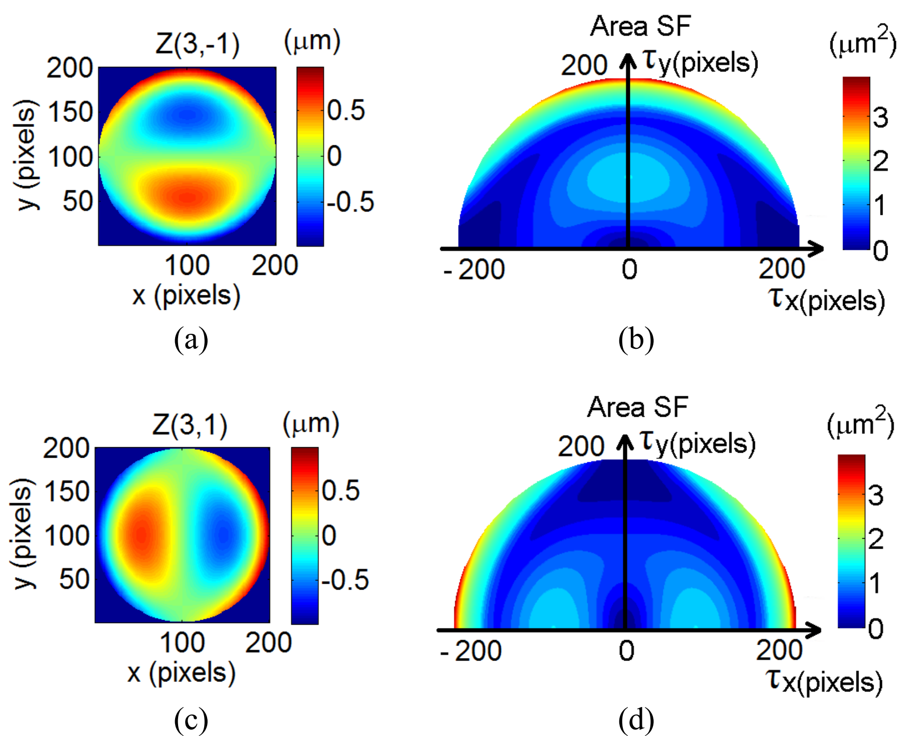


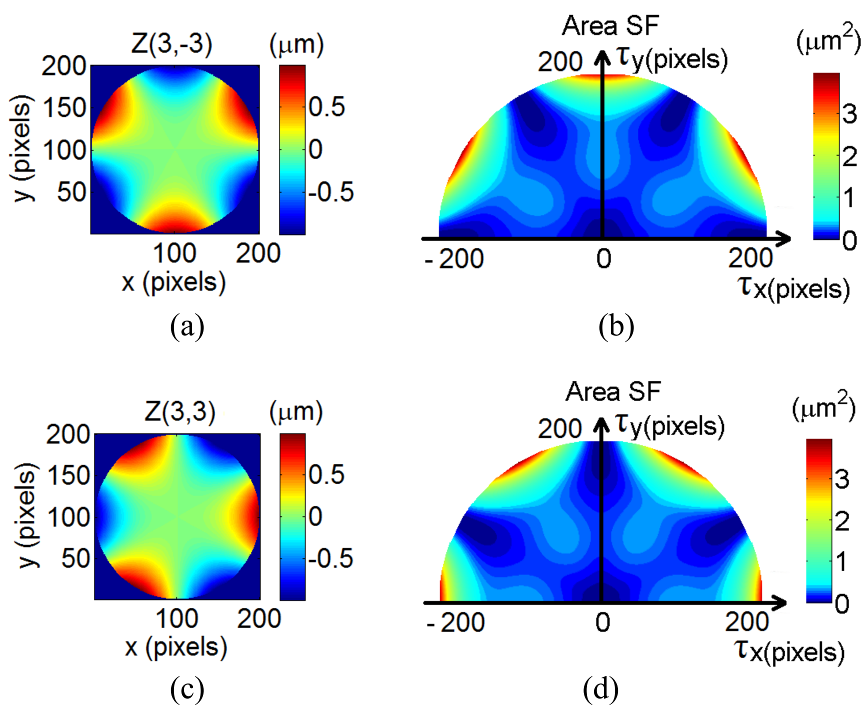
Figure 3.10 Area SF of astigmatism.



(a) (b)

(c) (d)

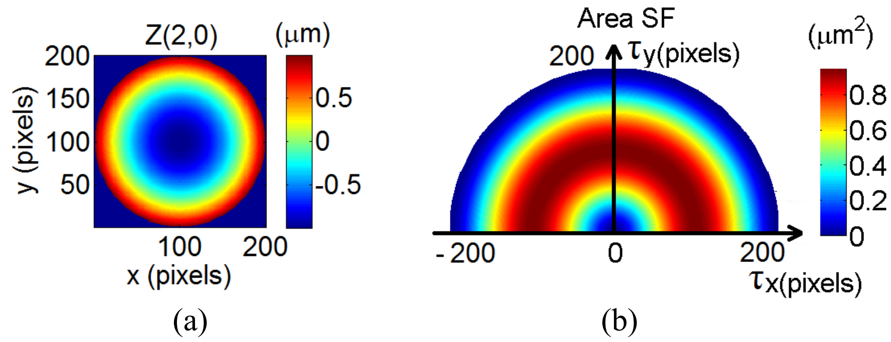
Figure 3.11 Area SF of coma.



(a) (b)

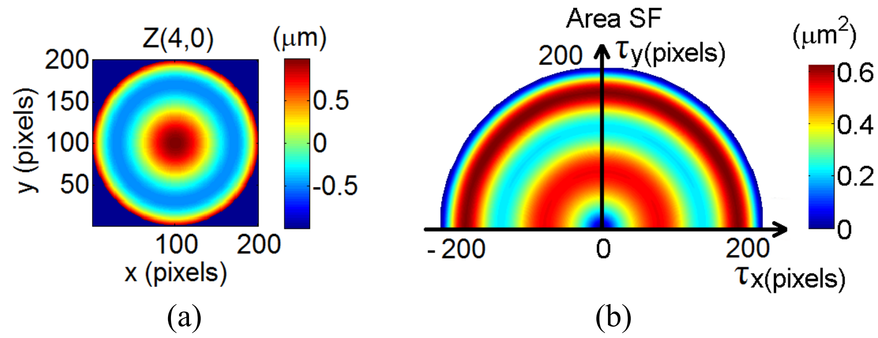
(c) (d)

Figure 3.12 Area SF of trefoil.



(a) (b)

Figure 3.13 Area SF of defocus.



(a) (b)

Figure 3.14 Area SF of third order spherical.

3.1.3.2 Total Area SF of a Surface Comprised of Zernike Polynomials

Similar to the linear SF analysis above, assume that the surface error is composed of two different Zernike polynomials in Eq. (3.19) and Eq. (3.20). Then the area SF is given as

$$S(\tau,\varphi)=E\left\{\left\{Z\left[(x,y)+(\tau\cos\varphi,\tau\sin\varphi)\right]-Z(x,y)\right\}^2\right\}=E\left\{\left\{Z(P+\tau(\varphi))-Z(P)\right\}^2\right\}, \quad (3.36)$$

where $Z(P+\tau(\varphi))$ is the surface height at a radial distance τ from the position P in a direction φ.

$$\begin{aligned} S(\tau,\varphi) &= E\left\{\{[Z_1(P+\tau(\varphi))+Z_2(P+\tau(\varphi))]-[Z_1(P)+Z_2(P)]\}^2\right\} \\ &\ldots\ldots \\ &= S_1(\tau,\varphi)+S_2(\tau,\varphi)+E\{Z_1(P+\tau(\varphi))Z_2(P+\tau(\varphi))\}+E\{Z_1(P)Z_2(P)\} \\ &\quad -E\{Z_1(P+\tau(\varphi))Z_2(P)\}-E\{Z_1(P)Z_2(P+\tau(\varphi))\}. \end{aligned} \quad (3.37)$$

In contrast to linear SF, each area SF $S(\tau,\varphi)$ corresponds to only one direction φ. Therefore, the integration from 0 to 2π is not applied, which means the cross terms cannot be cancelled. In short, even if the Zernike terms have different azimuthal frequencies, the area SF of the sum of the Zernike terms may not equal to the sum of the area SF of each Zernike term.

Figure 3.15, Figure 3.16, and Figure 3.17 give a demonstration of the total area SF of two Zernike terms with different azimuthal frequencies ($Z(2,2)$ and $Z(4,4)$). The result shows that the area SF of the total Zernikes is not equal to the sum of two area SFs.

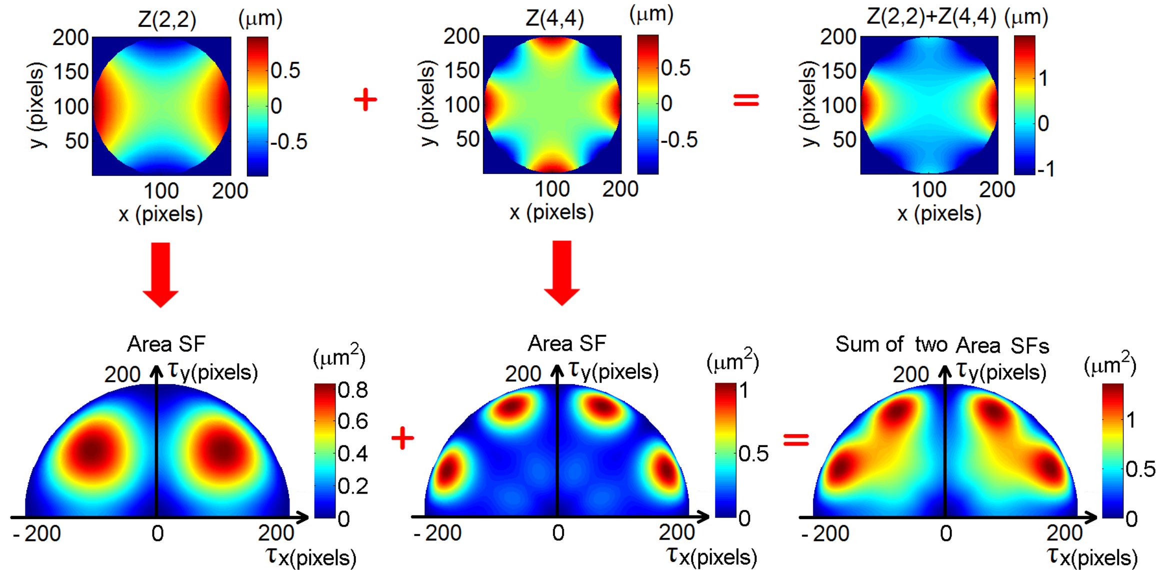


Figure 3.15 Sum of two area SFs of each of the Zernike polynomials with different azimuthal frequencies.

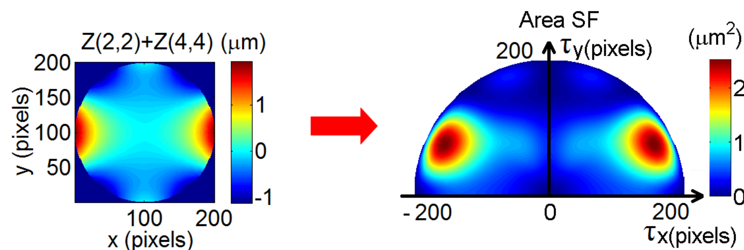


Figure 3.16 Area SF of the total Zernikes.

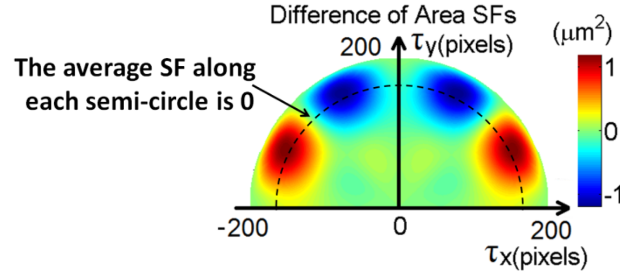


Figure 3.17 Difference between the SF of the total Zernikes and the sum of the two SFs.

However, as Figure 3.17 shows, if we integrate the area SF along the same radius, the average SF will be 0. In other words, the deviation of the linear SF for each radius τ is 0, which agrees well with the result for the total linear SF in Eq. (3.31).

3.2 Structure Function and Autocorrelation Function (ACF)

As noted in Chapter 1, ACF is an important tool for surface characterization. The concept of correlation is well known in statistics, where it is used to analyze the influence between two sets of numbers. Typically, the ACF is used to investigate the repeatability of a profile or a surface. This section discusses the relationship between the SF and the ACF.

3.2.1 Linear SF and Linear ACF

The linear SF can be described as

$$\begin{aligned} SF(\tau) &= E\left\{\left\{\left[z(x)+h\right]-\left[z(x+\tau)+h\right]\right\}^2\right\} = E\left\{\left[z(x)-z(x+\tau)\right]^2\right\} \\ &= E\left\{z^2(x)\right\}+E\left\{z^2(x+\tau)\right\}-2E\left\{z(x)z(x+\tau)\right\}, \end{aligned} \tag{3.38}$$

where $E\{\ \}$indicates an expectation, h is the mean of the surface, $z(x)$ and $z(x+\tau)$ are the surface heights after removing the mean value (piston). The function $E\{z(x)z(x+\tau)\}$ is the autocovariance function $ACV(\tau)$.

For a stationary surface [25] that has the same sub-aperture statistical characteristics,

$$E\left\{z^2(x)\right\} = E\left\{z^2(x+\tau)\right\} = ACV(0) = \sigma^2, \tag{3.39}$$

where σ^2 is the mean square of the surface after removing the mean value. Therefore,

$$SF(\tau) = 2\left\{\sigma^2 - ACV(\tau)\right\}. \quad (3.40)$$

Considering the relationship between ACV and ACF in Eq. (1.14), the expression of the linear SF can be rewritten as

$$SF(\tau) = 2\left\{\sigma^2 - \sigma^2 ACF(\tau)\right\} = 2\sigma^2\left\{1 - ACF(\tau)\right\}. \quad (3.41)$$

The ACF peaks at 1 for $\tau = 0$ and shows local maxima for τ length scale for which the surface has periodicity.

Thus, the linear SF is closely related to the linear ACF for a stationary surface. The overall scales of the two functions are different, and the locations of the maxima and minima are inverted.

3.2.2 Area SF and Area ACF

3.2.2.1 Stationary Surface

Similar to linear SF, the area SF can be described as,

$$\begin{aligned} SF(\tau_x, \tau_y) &= E\left\{\left\{\left[z(x,y)+h\right]-\left[z(x+\tau_x, y+\tau_y)+h\right]\right\}^2\right\} = E\left\{\left[z(x,y)-z(x+\tau_x, y+\tau_y)\right]^2\right\} \\ &= E\left\{z^2(x,y)\right\} + E\left\{z^2(x+\tau_x, y+\tau_y)\right\} - 2E\left\{z(x,y)z(x+\tau_x, y+\tau_y)\right\}, \end{aligned} \quad (3.42)$$

where the function $E\left\{z(x,y)z(x+\tau_x, y+\tau_y)\right\}$ is the area autocovariance function $ACV(\tau_x, \tau_y)$.

For a stationary surface that has the same sub-aperture statistical characteristics,

$$E\left\{z^2(x,y)\right\} = E\left\{z^2(x+\tau_x, y+\tau_y)\right\} = ACV(0,0) = \sigma^2. \quad (3.43)$$

Therefore,

$$SF(\tau_x,\tau_y) = 2\left\{\sigma^2 - ACV(\tau_x,\tau_y)\right\} = 2\left\{\sigma^2 - ACF(\tau_x,\tau_y)\sigma^2\right\} = 2\sigma^2\left\{1 - ACF(\tau_x,\tau_y)\right\}. \tag{3.44}$$

The ACF peaks at 1 for $\tau_x = \tau_y = 0$ and shows local maxima for τ_x and τ_y length scale for which the surface has periodicity.

Thus, the area SF is closely related to the ACF for a stationary surface. The overall scales of the two functions are different, and the locations of the maxima and minima are inverted.

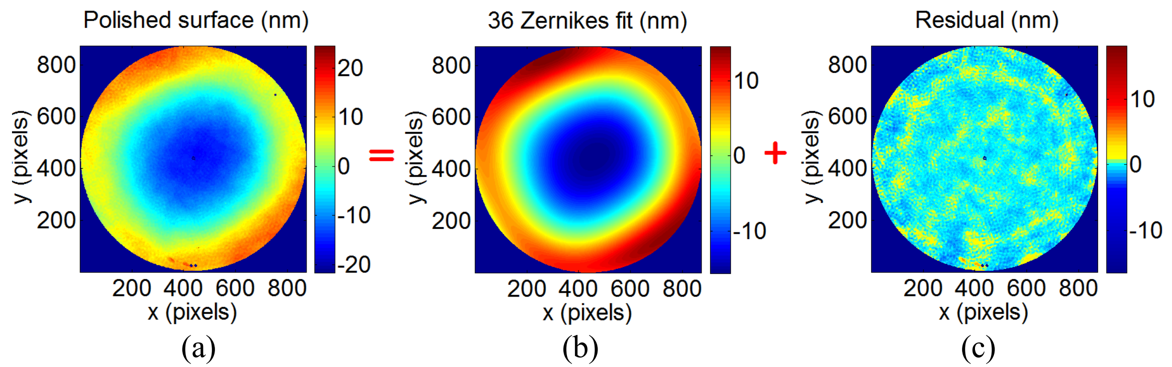


Figure 3.18 Zernikes fit of a polished surface. (a) Polished surface. (b) 36 Zernikes fit. (c) Residual surface error after 36 Zernikes removal.

Figure 3.18 demonstrates the 36 Zernikes fit of a conventionally polished flat. As Figure 3.18 (c) shows, the residual surface error is, typically, stationary and characterized by a speckled or "orange peel" texture. Figure 3.19 (a) and (b) are the area SF and area ACF for the residual error in this large polished flat. Figure 3.19 (c) is the difference between SF and $2\sigma^2\{1-ACF\}$, where the $\sigma^2 = 0.3966\,\text{nm}^2$. As expected, most of the differences are zero except at the edges where there are only a few calculated points.

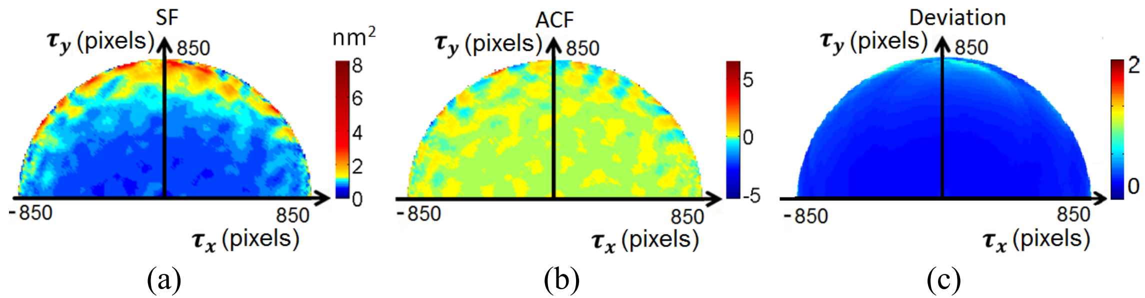


Figure 3.19 Relationship between area SF and area ACF for a polished surface after removing 36 Zernikes. (a) Area SF of the residual error. (b) Area ACF of the residual error. (c) Deviation=SF-2σ^2(1-ACF).

3.2.2.2 Non-stationary Surface

Optical surfaces, particular when figure is considered, are typically not stationary because they vary with translation.

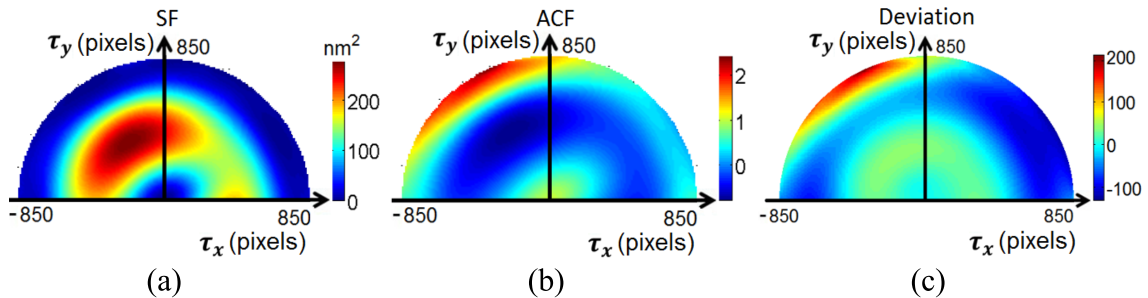


Figure 3.20 Relationship between area SF and area ACF for a non-stationary surface.(a) Area SF of the polished surface in Figure 3.18 (a). (b) Area ACF of the same polished surface. (c) Deviation=SF-2σ^2(1-ACF).

In this non-stationary case, Eq. (3.44) is not correct. For example, Figure 3.20 shows the relationship between the area SF and area ACF for the polished surface in Figure 3.18 (a). As Figure 3.20 (c) denotes, SF-2σ^2(1-ACF)$\neq$0. Hence the simple relationship between SF and ACF does not apply to non-stationary surface.

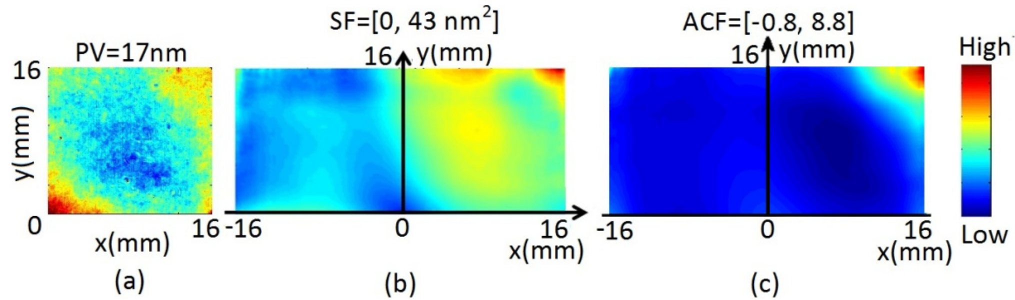


Figure 3.21 Analysis of a polished-silicon flat.(a) Input map from Fizeau interferometer. (b) Area SF. (c) Area ACF.

Further insight into the difference between the area SF and the area ACF is obtained from an analysis of a polished-silicon flat measured with a Fizeau interferometer (Figure 3.21). The area SF calculation is shown in Figure 3.21 (b). The astigmatic form in the surface leads to larger SF values in the first quadrant compared to the second. In Figure 3.21 (c), the area ACF contains similar information as area SF, but it is not as distinct as area SF.

Comparison between Figure 3.21 (b) and (c) indicates that the area SF provides a superior visualization of the spatial content of the surface as a function of separation and orientation. As Sayles and Thomas [25] point out, the area ACF is the sum of terms each of which is the product of two amplitudes, which is difficult to understand and not intuitive. From the definitions of SF and ACF, it is obvious that the area SF is independent of the mean plane, while the area ACF is strongly influenced by the mean plane. Moreover, because the area ACF is based on the product of two amplitudes and the area SF is based on the difference of two amplitudes, the computation speed of the area ACF is slower than area SF, especially for big arrays.

3.3 Structure Function and Power Spectral Density (PSD) Function

Currently, PSD is commonly used for describing the spatial frequency characteristics. It is the square of the Fourier transform of the surface or profile, with appropriate normalization [48,91]. However, it has some disadvantages. This section discusses the relationship between the PSD and SF, and makes comparisons between these two representations.

3.3.1 Linear SF and Linear PSD

The linear SF is usually estimated over the whole surface, while the linear PSD is normally calculated along the profiles [24,91].

3.3.1.1 Relationship between Linear SF and Linear PSD for Stationary Process

(1) Linear PSD and Linear ACF

In addition to the traditional calculation from the surface height data (introduced in Chapter 1), the linear PSD can also be evaluated from the ACF [92,93]. This method is based on the Wiener–Khinchin theorem [94] which states that the ACF of a well behaved stationary random process has a spectral decomposition defined by the power spectrum of that process:

$$P(f_m) = \sum_{n=-N/2}^{N/2} ACV(n) e^{-i2\pi mn/N}, \tag{3.45}$$

where $P(f_m)$ is the description of PSD used in Wiener–Khinchin theorem:

$$P(f_m) = \frac{1}{N} \left| DFT(f_m) \right|^2 = \frac{1}{N} \left| \sum_{n=0}^{N-1} z(n) e^{-i2\pi mn/N} \right|^2. \tag{3.46}$$

Multiplying both sides of Eq. (3.45) with $\Delta\tau$, then

$$PSD(f_m) = \frac{\Delta\tau}{N} \left| \sum_{n=0}^{N-1} z(n) e^{-i2\pi mn/N} \right|^2 = \Delta\tau \sum_{n=-N/2}^{N/2} ACV(n) e^{-i2\pi mn/N}, \tag{3.47}$$

where $\Delta\tau$ is the sampling separation for the ACV and PSD, and it is usually equal to the sampling interval Δx of the profile measurements or a multiple of it.

Since the ACV is an even function, the equation can be rewritten as [46]

$$PSD\left(f_m\right) = 2\Delta\tau \sum_{n=0}^{N/2} ACV\left(n\Delta\tau\right)\cos\left(2\pi f_m n\Delta\tau\right). \qquad (3.48)$$

In short, the PSD is the product of $\Delta\tau$ and Fourier transform of the autocovariance function,

$$PSD(f) = \Delta\tau.\mathscr{F}\left[ACV(\tau)\right]. \qquad (3.49)$$

In order to verify the relationship between PSD and ACV, a sinusoidal profile (λ = 400 µm) was generated in Figure 3.22 (a). Then the PSD was calculated in two ways, respectively.

The one computed directly from the profile data is shown in Figure 3.22 (b). Note that the spatial frequency spike in PSD at 0.0025 μm^{-1} is consistent with the wavelength of the sinusoidal profile ($1 / 400\mu m = 0.0025\mu m^{-1}$).

The other one is computed from the ACV. The two-side linear ACV is shown in Figure 3.22 (c). As discussed in section 1.3.1.2, the ACF with shorter separation is more accurate due to the calculation with more points. Therefore, the ACF with separation shorter than 2000 µm in both sides is chosen for computing the PSD. As a result, this PSD is shown in Figure 3.22 (d).

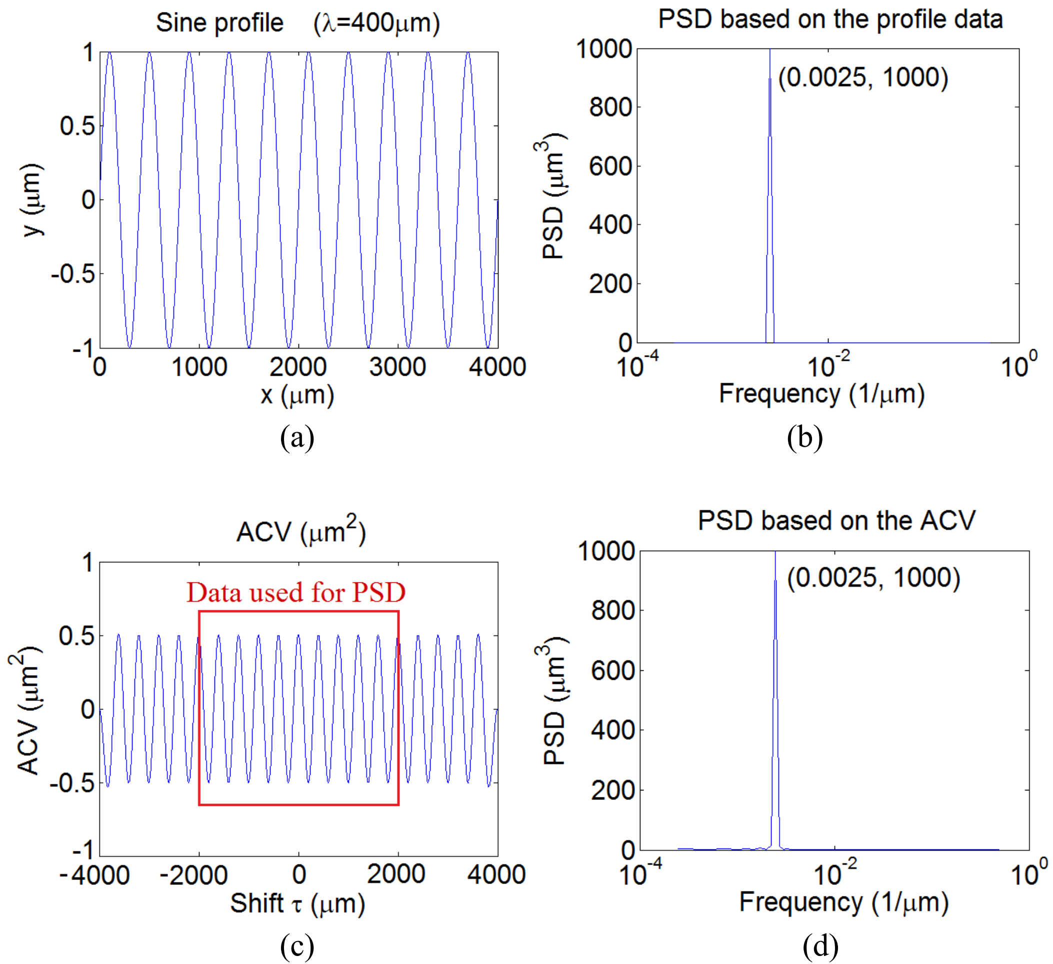


Figure 3.22 Comparison of PSD based on profile data and ACV. (a) A sinusoidal profile with λ= 400 μm. (b) The linear PSD calculated from the profile data. (c) The linear ACV of the profile data. (d) The linear PSD calculated from the ACV.

It is clear that the two PSDs are almost identical. They both exactly indicate the periodicity of the sinusoidal profile. However, there may be some insignificant difference because the length of the profile is finite, which can cause two main problems:

First, the points used for computing the ACV are different for different separation τ, which means the degree of accuracy of individual ACV value can vary. Consequently, the PSD based on the ACV is not perfect, because the statistical reliability of the information is limited.

Second, the finite length of the ACV may cause spurious edge effects for computing the PSD, because DFT always processes a profile as if it were part of a periodic function of identical curves extending to infinity. To solve this problem, it is important to multiply the data with a “window function” before applying the DFT. However, there is no strict standard for choosing the window. With different weighting windows, the PSD may be different. The details of the window function will be discussed in the following sections.

Taking all the factors mentioned above into account, the PSD evaluated from the ACV can meet the requirement of application to some extent. In most cases, its accuracy is good enough since the computing points are at least thousands.

Under ideal conditions, if the input profile were infinite, the PSD evaluated from the ACV would be identical to the PSD calculated directly from the profile data.

(2) Linear PSD and linear SF

Since the SF is related to ACV for stationary surface (Eq. (3.40)), the relationship between SF and PSD for a stationary surface can be shown as

$$PSD(f) = \Delta\tau \cdot \mathscr{F}[ACV(\tau)] = \Delta\tau \cdot \mathscr{F}[\sigma^2 - \frac{1}{2}SF(\tau)]. \quad (3.50)$$

In order to verify the relationship between the linear PSD and linear SF, a simulation is shown in Figure 3.23. Figure 3.23 (a) is the two-side linear SF calculated from the same profile in Figure 3.22 (a). And the central part of the linear SF was chosen for computing the PSD. It is clear that the PSD based on the SF is almost identical to the PSD evaluated directly from the profile data in Figure 3.22 (b). Similar to that based on the ACV, the PSD based on the SF also has the problems of statistical reliability and the spurious edge effects when using the FFT, but its accuracy is good enough in most cases.

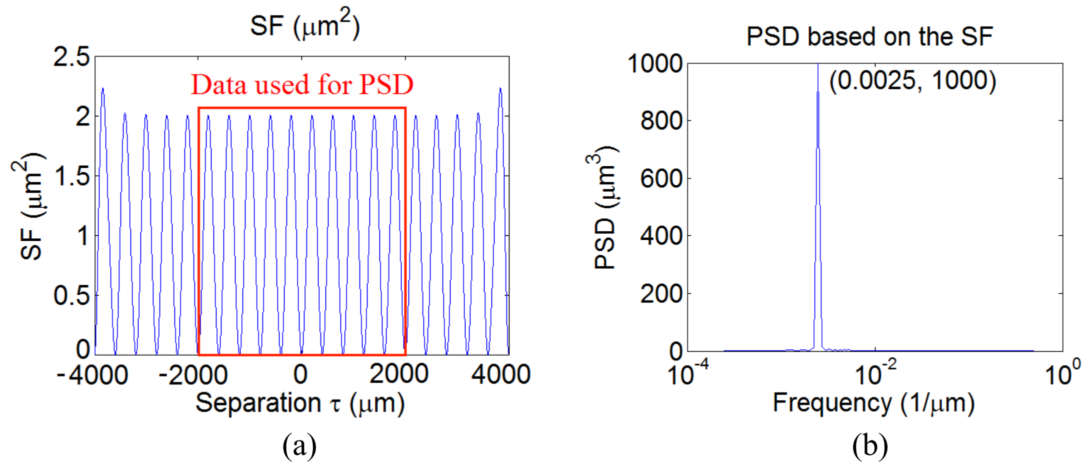


Figure 3.23 Calculation of the linear PSD based on the SF. (a) The linear SF of the sine profile in Figure 3.22 (a). (b) The linear PSD calculated from the SF.

3.3.1.2 Comparison between Linear SF and Linear PSD

(1) Linear PSD uses small fraction of surface data

To investigate the true linear PSD of a surface with random roughness, it is necessary to evaluate the ensemble average of some linear PSDs computed from profiles measured in different places of the surface [32]. Normally, analysis based on a single profile is insufficient because the curve may contain noise and lack reproducibility.

Usually, there are two ways to get the ensemble average of linear PSDs [32]:

The first way is to measure numbers of profiles in different places of the surface, compute the linear PSD for each profile, and average all the PSDs to a smooth and noise-free curve. The disadvantage of this method is that it is time consuming when measuring numbers of the profiles.

The other way is to use only one profile. When the spatial frequency needed in the PSD is much higher than the lowest spatial frequency evaluated from the profile length, the profile can be divided into small segments. After computing the PSD of each sub-profile and getting the average, it will lead to a smooth and noise-free curve.

Apparently, the problem of this method is that it loses the low spatial frequency components in the original profile.

If the surface comprises periodic or unidirectional structure instead of random or isotropic topography, the profiles should be measured in the parallel and perpendicular direction to the surface structure. If not, the analysis may contain spurious results.

In order to compare the linear PSD to the linear SF, we simulated a sinusoidal surface with λ= 400 μm (Figure 3.24).

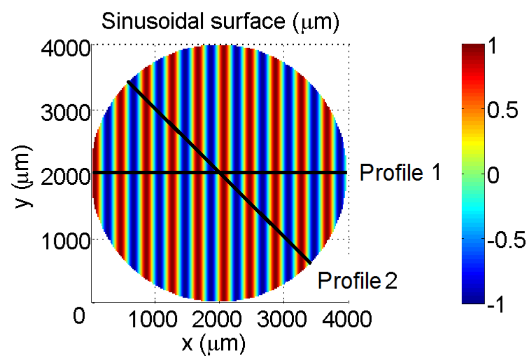


Figure 3.24 Simulated sinusoidal surface with λ= 400 μm. (Two profiles were picked for following analysis.)

Two profile were picked from the sinusoidal surface for analyzing the linear PSD. From Figure 3.25, it is clear that different profiles from the surface lead to different linear PSDs (both in the spatial frequency and magnitude).

In other words, since the linear PSD only uses a small fraction of the surface data, we need to choose the appropriate profiles that represent the true characteristics of the surface. For example, when a Talysurf is used, the measuring trace should be exactly parallel to the sinusoidal direction. However, it is very difficult to find the exact direction in practice.

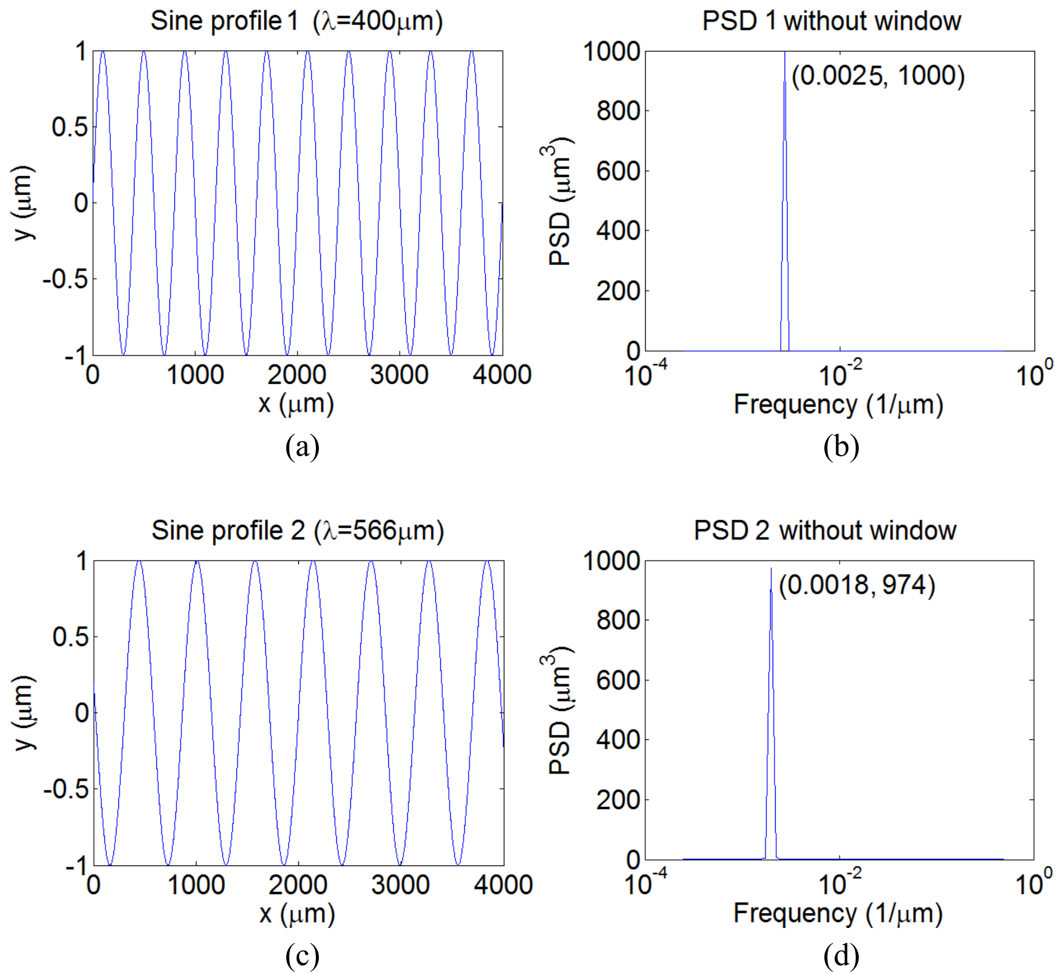
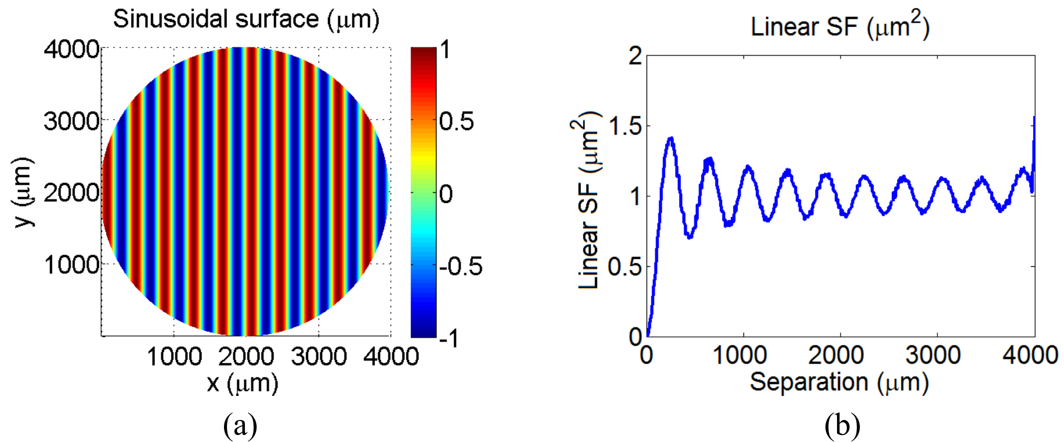


Figure 3.25 Calculation of linear PSD without window function. (a) Profile 1 from the sinusoidal surface. (b) Linear PSD of the profile 1 without window. (c) Profile 2 from the sinusoidal surface. (d) Linear PSD of the profile 2 without window.

The other way to calculate the linear PSD here is to integrate the area PSD in x or y axis. For instance, if we measure the sinusoidal surface with a Fizeau interferometer, the area PSD can be calculated from the surface data. After integrating the area PSD in x or y direction, a linear PSD can be obtained. Particularly, during the measurement, the coordinate of the Fizeau interferometer should be consistent with that of the surface structure. In this case, the x or y axis should match well with the sinusoidal direction. In brief, it is difficult for the operation.



(a) (b)

Figure 3.26 Linear SF of the sinusoidal surface. (a) Simulated sinusoidal surface with λ= 400 μm. (b) Linear SF of the whole sinusoidal surface.

Unlike the linear PSD, the calculation of linear SF does not have the problems mentioned above. All the measured surface height data can be used to estimate the linear SF, it is unnecessary to pick profiles for analysis. Therefore, the calculation of linear SF is explicit and the result is definite.

Moreover, the calculation of linear SF is not influenced by the coordinate system of the instrument, because the linear SF is based on the sampled paired-data in every direction. In this case, there is no need to match the x or y axis of the instrumental coordinate system to the sinusoidal direction in the surface.

In brief, a surface is definitely corresponding to only one linear SF, regardless of the coordinate systems.

(2) Linear PSD needs window functions

During the calculation of PSD, the discrete Fourier transform (DFT) always processes a profile as if it were part of a periodic function of identical curves extending to infinity. In other words, there are strong edge effects between the neighbors of such a periodic function. The discontinuity at edges spreads power all across the spectrum. To solve this

problem, it is important to multiply the data with a “window function” before applying the DFT. The window function starts around zero, then increases to a maximum at the center of the image and decreases again [42,95]. Thus the discontinuity can be corrected.

Therefore, the one-side linear PSD with window can be expressed as

$$PSD(f_m) = \frac{2\Delta x}{N}\left|\sum_{n=0}^{N-1} e^{-i2\pi mn/N} z(n)w(n)\right|^2 K(m), \qquad (0 \le m \le \frac{N}{2}) \tag{3.51}$$

where $w(n)$ is the chosen window function.

Under this condition, the values of $PSD(f_m)$ are changed because of the application of $w(n)$. Therefore, Eq. (1.4) is not satisfied, thus, $RMS^2 \ne f_1 \sum_{m=0}^{N/2} PSD(f_m)$.

To solve this problem, a scaling factor S is adopted:

$$PSD_{normal}(f_m) = \frac{PSD(f_m)}{S} = \frac{2\Delta x}{SN}\left|\sum_{n=0}^{N-1} e^{-i2\pi mn/N} z(n)w(n)\right|^2 K(m). \quad (0 \le m \le \frac{N}{2}) \tag{3.52}$$

According to Eq. (1.4)

$$\sigma^2_{windowed} = f_1 \sum_{m=0}^{N/2} PSD(f_m) = S f_1 \sum_{m=0}^{N/2} PSD_{normal}(f_m) = S\sigma^2_{original}, \tag{3.53}$$

where $\sigma^2_{windowed}$ is the mean square of the windowed data $z(n)\, w(n)$, and $\sigma^2_{original}$ is the mean square of the original data $z(n)$. Thus, the scaling factor S is obtained:

$$S = \frac{\sigma^2_{windowed}}{\sigma^2_{original}}. \tag{3.54}$$

Therefore, after the normalization, the first definition of linear PSD is

$$\begin{aligned} PSD_{normal}(f_m) &= \frac{\sigma^2_{original}}{\sigma^2_{windowed}} PSD(f_m) \\ &= \frac{2\Delta x \sigma^2_{original}}{N\sigma^2_{windowed}}\left|\sum_{n=0}^{N-1} e^{-i2\pi mn/N} z(n)w(n)\right|^2 K(m). \qquad (0 \le m \le \frac{N}{2}) \end{aligned} \tag{3.55}$$

Thus, the area under this normalized PSD curve is equal to the RMS^2 of the original profile data.

Similarly, the two-side linear PSD with window can be expressed as

$$PSD(f_m) = \frac{\Delta x}{N} \left| \sum_{n=0}^{N-1} e^{-i2\pi mn/N} z(n) w(n) \right|^2 K(m). \qquad (-\frac{N}{2} \le m \le \frac{N}{2}) \qquad (3.56)$$

Its normalized expression is

$$\begin{aligned} PSD_{normal}(f_m) &= \frac{\sigma^2_{original}}{\sigma^2_{windowed}} PSD(f_m) \\ &= \frac{\Delta x \sigma^2_{original}}{N \sigma^2_{windowed}} \left| \sum_{n=0}^{N-1} e^{-i2\pi mn/N} z(n) w(n) \right|^2 K(m). \qquad (-\frac{N}{2} \le m \le \frac{N}{2}) \end{aligned} \qquad (3.57)$$

The third definition of linear PSD (based on ACV) with window can be expressed as

$$PSD(f_m) = \Delta\tau \sum_{n=-N/2}^{N/2} ACV(n) w(n) e^{-i2\pi mn/N}. \qquad (3.58)$$

The related normalization is

$$PSD_{normal}(f_m) = \frac{\sigma^2_{original}}{\sigma^2_{windowed}} PSD(f_m) = \frac{\Delta\tau\sigma^2_{original}}{\sigma^2_{windowed}} \sum_{n=-N/2}^{N/2} ACV(n) w(n) e^{-i2\pi mn/N}. \qquad (3.59)$$

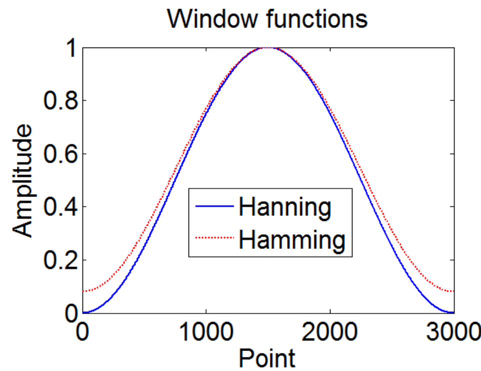


Figure 3.27 Hanning and Hamming windows in spatial domain.

Taking the Hanning window as an example (Figure 3.27), which is defined as,

$$w_i = \frac{1}{2}[1 - \cos(\frac{2\pi i}{N})]. \qquad (i = 0, 1, \ldots\ N-1) \qquad (3.60)$$

It has the following symmetry,

$$w_i = w_{N-i} \, . \tag{3.61}$$

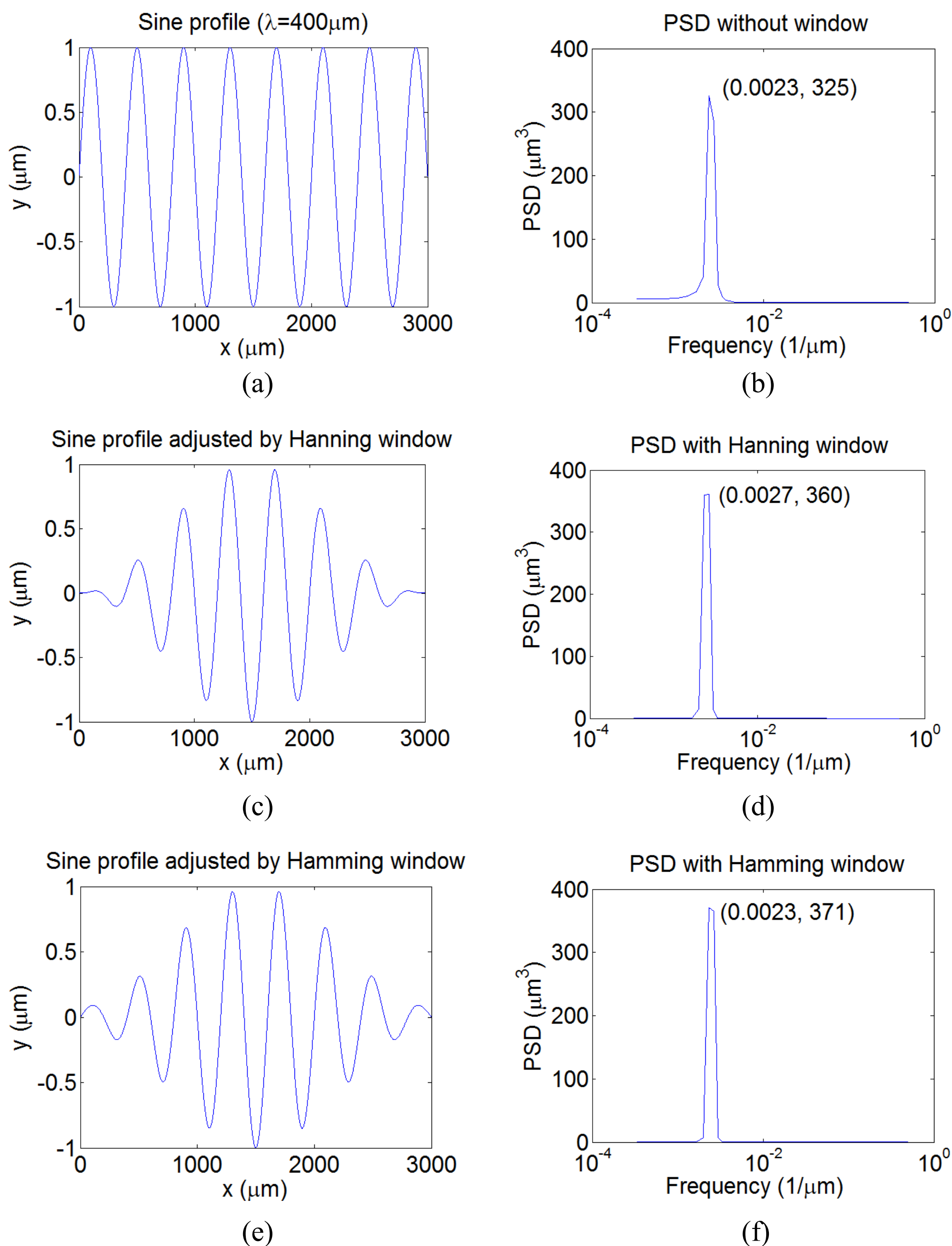


Figure 3.28 Calculation of linear PSD in different ways. (a) Sine profile with λ= 400 μm. (b) Linear PSD without window. (c) Sine profile adjusted by Hanning window. (d) Linear PSD with Hanning window. (e) Sine profile adjusted by Hamming window. (f) Linear PSD with Hamming window.

Figure 3.28 (a) is a profile from the sinusoidal surface in Figure 3.24. This sinusoidal profile contains 7.5 periods ($\lambda = 400\ \mu m$). As shown in Figure 3.28 (a), when using the DFT to extend both sides of the sinusoidal function, there should be strong spurious edge effects, because 7.5 periods do not cover the integrated periodic phase. In other words, if the profile length satisfies $l = k\lambda$ (k = 1, 2, 3...), the edge effects can be eliminated. Thus, the linear PSD (without using window) shown in Figure 3.28 (b) is not accurate.

To reduce to edge effects, a Hanning window is applied in Figure 3.28 (c), and the windowed PSD is shown in Figure 3.28 (d). It is clear that both spiky frequency and magnitude are different from the PSD without using window.

Moreover, a Hamming window is also applied in Figure 3.28 (e), and the windowed PSD is shown in Figure 3.28 (f). It is obvious that the result is different from the PSD with Hanning window.

In fact, there are various windows, which may lead to different results. Thus, the linear PSD is strongly dependent on the choice of windows. Unfortunately, there is no strict standard for choosing the window. The best thing we can do is to use a window that has a highly concentrated central lobe and small side lobes with high decaying speed [96].

(3) Low-order terms removal

Figure 3.29 shows the combination of a simulated radial sinusoidal surface and a simulated defocus. Compared to the sinusoidal surface with 10 periods, the defocus can be taken as a low-order form.

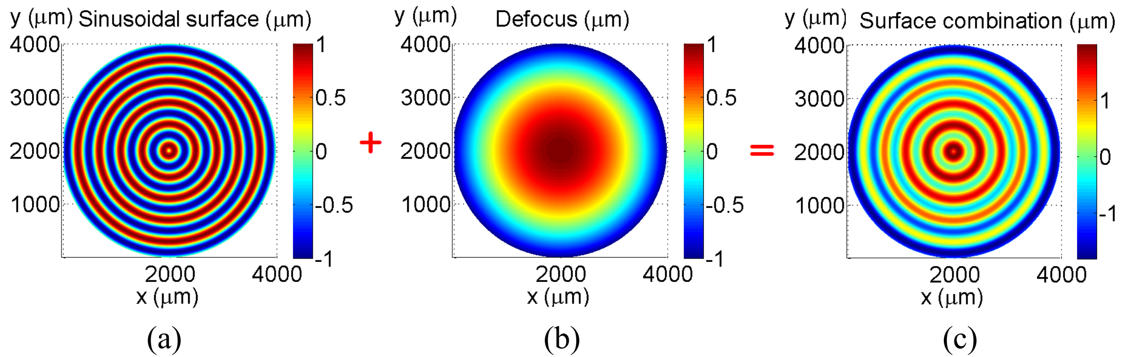


Figure 3.29 Combination of two surfaces. (a) Simulated sinusoidal surface with λ= 400 μm. (b) Defocus. (c) Combined surface.

In order to investigate the spatial content with linear PSD, it is necessary to choose some profiles from the surface. Since it contains a radial sinusoidal surface, a diametric profile is chosen in Figure 3.30 (a). However, as Figure 3.30 (b) shows, it is not an exact sinusoidal profile because the discontinuity of the central part. Thus, we choose a radial profile instead in Figure 3.30 (c). As Figure 3.30 (d) shows, it contains 5 periods of the sinusoidal function. After adjusting with a Hanning window, the linear PSD is shown in Figure 3.30 (f). The spiky spatial frequency of 0.0025 μm^{-1} is consistent with the period of the sinusoidal function (1/400=0.0025). The frequency of 0.0005 μm^{-1} also corresponds to a spike, which means the surface contains a low-order term with period of 2000 μm (1/2000=0.0005). Apparently, this is unreasonable because the diameter of the defocus is 4000 μm. Moreover, the low-order terms in the profile may reduce the accuracy of computing other spatial frequency components. Consequently, the low-order term should be removed before the calculation of the linear PSD [41,42,92,93].

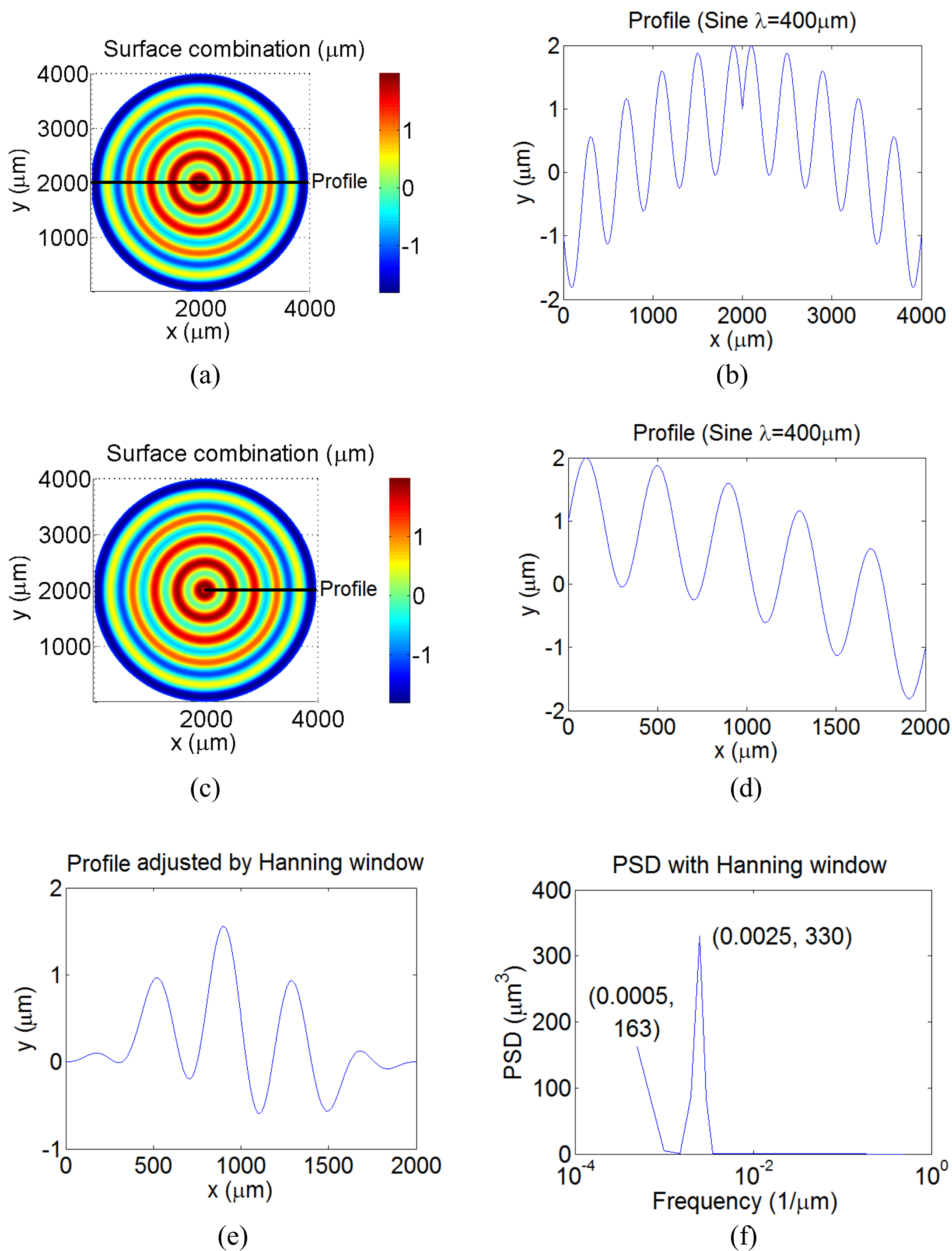
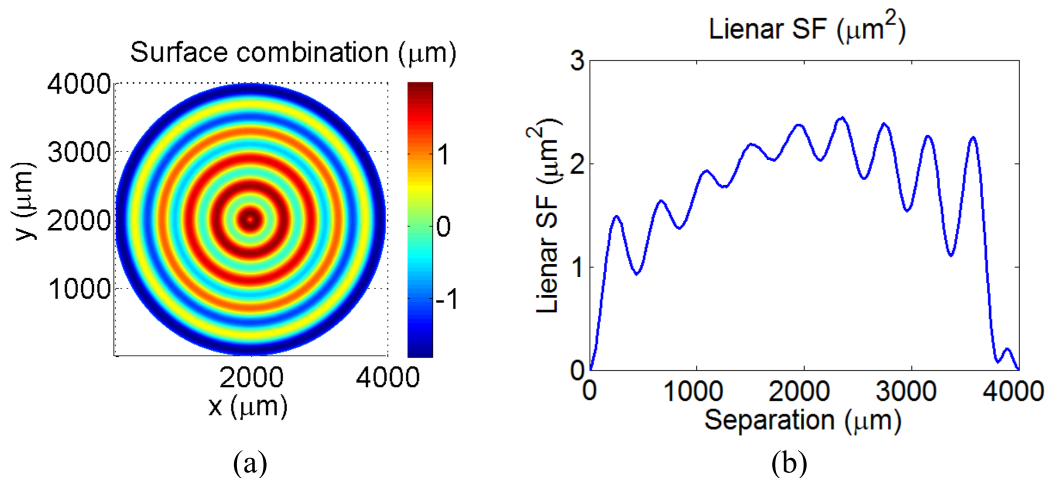


Figure 3.30 Calculation of linear PSD with Hanning window. (a) Combined surface. (b) Diametric profile from the combined surface. (c) Combined surface. (d) Radial profile from the combined surface. (e) The radial profile adjusted by Hanning window. (f) Linear PSD of the adjusted radial profile.

For the linear SF, it is not necessary to remove the low-order term before the calculation. Figure 3.31 shows a linear SF of the whole combined surface. It reflects the trend of the defocus and keeps the high spatial frequency components. In this rotationally invariant surface, the uneven value of the peaks is because the linear SF is calculated from every possible paired data. If the paired data are not in the radial direction, the related SF value will not represent the true spatial component.



(a) (b)

Figure 3.31 Linear SF of the combined surface. (a) Combined surface. (b) Linear SF of the combined surface.

In fact, if the surface is rotationally invariant, the analysis with a radial profile represents all the characteristics. In this case, we picked the same radial profile as that for computing linear PSD, and calculated the linear SF in Figure 3.32 (b). It is clear that the linear SF indicates both the high spatial frequency and low-order trend.

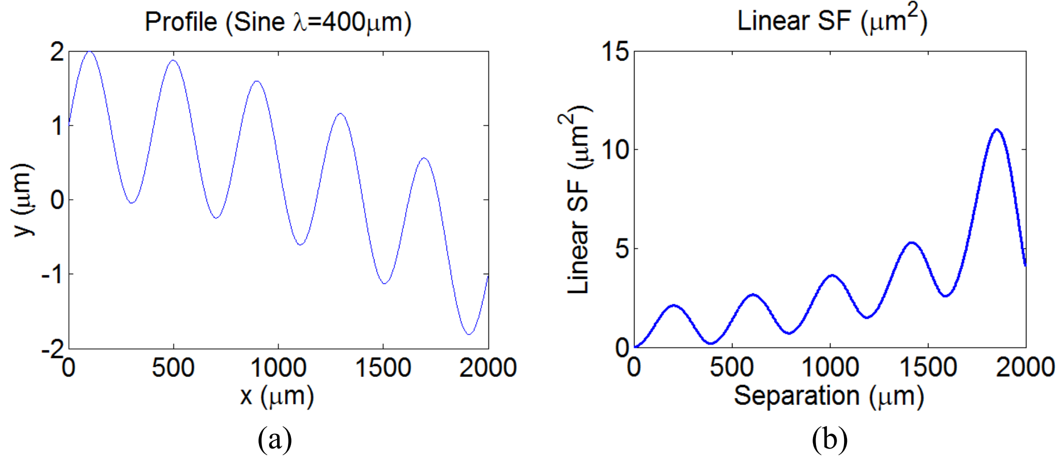


Figure 3.32 Linear SF of the combined profile. (a) Radial profile from the combined surface. (b) Linear SF of the profile.

3.3.2 Area SF and Area PSD

With the development of 3-D surface measuring instruments, researchers have extended the linear analysis to areal investigations in order to obtain a better and comprehensive representation of the surface. This section discusses the relationship between the area PSD and area SF, and makes comparisons between these two representations.

3.3.2.1 Relationship between Area PSD and Area SF for Stationary Process

Similar to the linear analysis, it is possible to connect the area PSD to the area ACV based on the Wiener–Khinchin theorem:

$$P(f_x,f_y)=\sum_{n_x=-N_x/2}^{N_x/2}\sum_{n_y=-N_y/2}^{N_y/2} e^{-i2\pi\left(f_x n_x \Delta x+f_y n_y \Delta y\right)} ACV(n_x,n_y), \qquad (3.62)$$

where $P(f_x,f_y)$ is the description of PSD used in Wiener–Khinchin theorem:

$$P(f_x,f_y)=\frac{1}{N_x N_y}\left|DFT(f_x,f_y)\right|^2=\frac{1}{N_x N_y}\left|\sum_{n_x=0}^{N_x-1}\sum_{n_y=0}^{N_y-1} z(n_x,n_y)e^{-i2\pi\left(f_x n_x \Delta x+f_y n_y \Delta y\right)}\right|^2 \qquad (3.63)$$

Multiplying both sides of Eq. (3.63) with $\Delta x \Delta y$, then

$$
\begin{aligned}
PSD(f_x, f_y) &= \Delta x \Delta y P(f_x, f_y) = \frac{\Delta x \Delta y}{N_x N_y} \left| \sum_{n_x=0}^{N_x-1} \sum_{n_y=0}^{N_y-1} z(n_x, n_y) e^{-i2\pi\left(f_x n_x \Delta x + f_y n_y \Delta y\right)} \right|^2 \\
&= \Delta x \Delta y \sum_{n_x=-N_x/2}^{N_x/2} \sum_{n_y=-N_y/2}^{N_y/2} ACV(n_x, n_y) e^{-i2\pi\left(f_x n_x \Delta x + f_y n_y \Delta y\right)}
\end{aligned} \qquad (3.64)
$$

where Δx and Δy are the sampling separation for the ACV and PSD, and they are usually equal to the sampling interval of the surface measurements in x and y dirctions or a multiple of them.

Since the ACV is an even function, the equation can be rewritten as

$$
PSD\left(f_x, f_y\right) = 2\Delta x \Delta y \sum_{n_x=0}^{N_x/2} \sum_{n_y=0}^{N_y/2} ACV(n_x \Delta x, n_y \Delta y) \cos\left[2\pi\left(f_x n_x \Delta x + f_y n_y \Delta y\right)\right]. \qquad (3.65)
$$

In short, the PSD is the product of $\Delta x \Delta y$ and Fourier transform of the autocovariance function,

$$
PSD(f_x, f_y) = \Delta x \Delta y \cdot \mathscr{F}\left[ACV(\tau_x, \tau_y)\right]. \qquad (3.66)
$$

Therefore, the relationship between area SF and area PSD for a stationary surface can be shown as

$$
PSD(f_x, f_y) = \Delta x \Delta y \cdot \mathscr{F}\left[\sigma^2 - \frac{1}{2} SF(\tau_x, \tau_y)\right]. \qquad (3.67)
$$

3.3.2.2 Comparison between Area SF and Area PSD

(1) Window function

The same as the linear PSD, it is also necessary to use the window functions for calculating the area PSD.

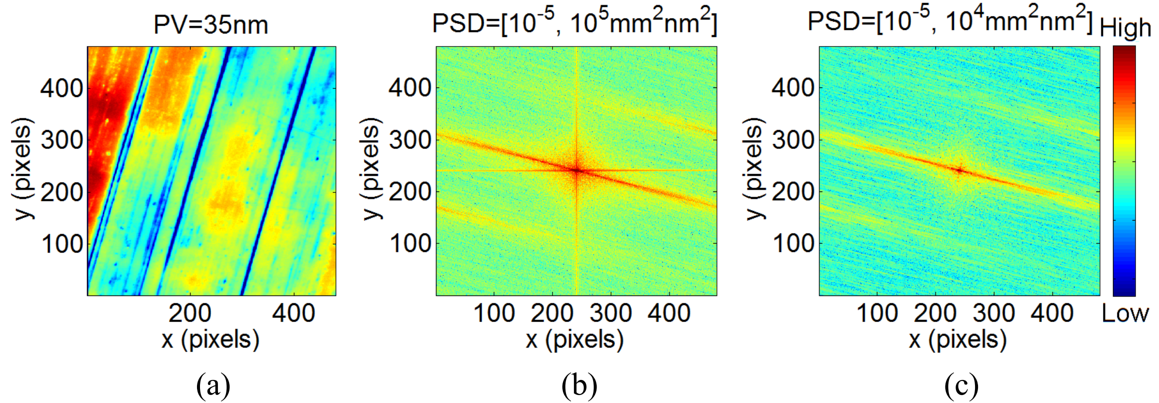


Figure 3.33 Analysis of the use of window. (a) Diamond-turned aluminum flat. (b) Area PSD of the flat without using window. (c) Area PSD with Hanning window.

Figure 3.33 (a) shows a diamond-turned aluminum flat measured with a scanning white light interferometer (SWLI). Figure 3.33 (b) shows its area PSD without applying any windows. It is clear that there are strong spurious edge effects along the x and y spatial frequency axes. The DFT always processes an image as if it were part of a periodic matrix of identical images extending to infinity. However, for this input height map, both x and y edges cannot match their opposite edges to be a smooth surface. Consequently, this discontinuity leads to the spurious power on the axes.

After applying the Hanning window, as shown in Figure 3.33 (c), the edge effects are eliminated. Because the window function starts around zero, then increases to a maximum at the center of the image and decreases again, therefore the discontinuity can be corrected.

In a word, the PSD with window function provides a better estimate of the surface structure. However, as discussed in Figure 3.28, there are various windows, which may lead to different results. As a result, the area PSD is strongly dependent on the choice of windows. Unfortunately, there is no strict standard for choosing the window.

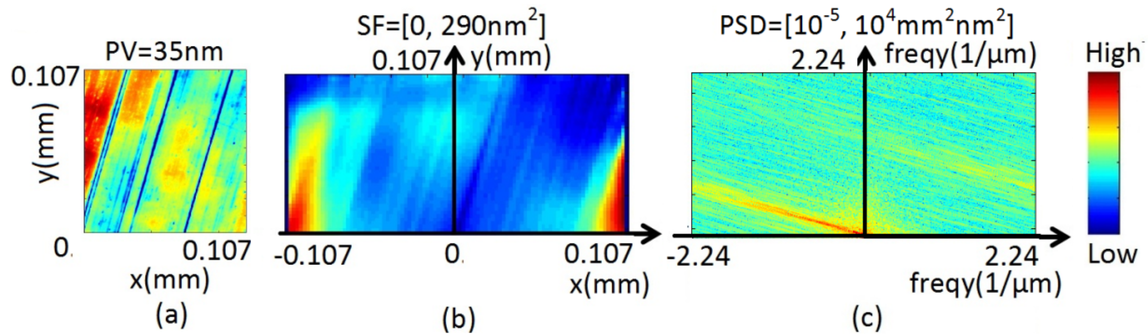


Figure 3.34 Analysis of a diamond-turned aluminum flat. (a) Input map from a SWLI measurement. (b) Area SF. (c) Area PSD.

Figure 3.34 (a) shows an analysis of the diamond-turned aluminum flat with area SF and PSD. Figure 3.34 (b) shows the area SF, which contains all the information of linear scratches of the surface. Figure 3.34 (c) shows the area PSD calculation with the Hanning window. In order to compare to the area SF, only two quadrants of the area PSD are demonstrated. The third and fourth quadrants contain the same information as the first and second quadrants. The area PSD characterizes the scratch information, but not as intuitively as the area SF.

In short, the area SF calculation is specific and correct, because it uses all the data and does not need any window functions. For the area PSD, the calculation varies with the choice of windows.

(2) Zero padding and low-order terms removal

Figure 3.35 shows the simulated surface map that is generated with a single Zernike term Z(2,-2), as well as its area SF and area PSD. Figure 3.35 (b) shows that the area SF captures the long spatial wavelength detail in a surface, while Figure 3.35 (c) exemplifies that the area PSD is inappropriate for representation of low order figure errors. Since the

2D DFT cannot cope well with the circular aperture shape, it is necessary to pad the circular shape to be rectangular.

In brief, the result of area PSD is also strongly dependent on zero padding and low-order terms removal.

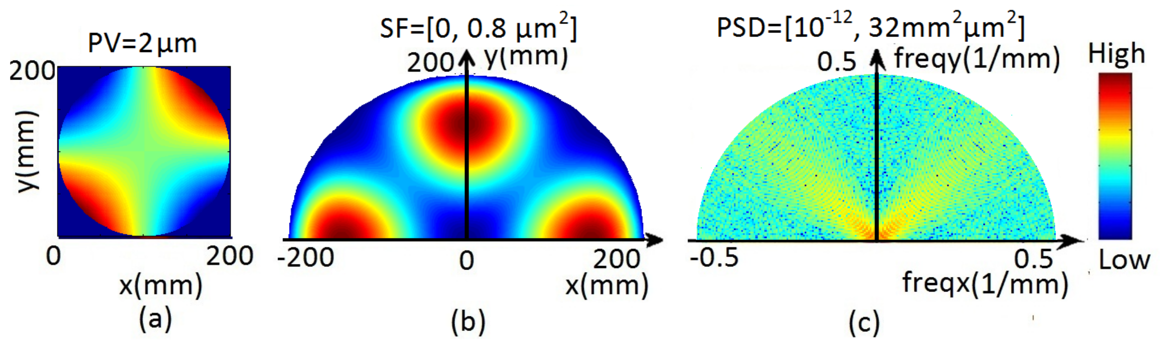


Figure 3.35 Analysis of simulation data.(a) Input map. (b) Area SF. (c) Area PSD.

(3) Arbitrary aperture shape

A key advantage of the area SF is the computational correctness for any arbitrary aperture shape. Figure 3.36 (a) shows a simulated hexagonal shape, representing a segment of a primary mirror of a large telescope. It is easy to calculate the area SF with all the data. To calculate an area PSD, it is necessary first to remove the form. In addition, it is difficult to calculate the area PSD, because we need to mask the full data aperture and separate it into rectangular sub-apertures, which should be "representative" of the true surface characteristics [43]. But for many optical surfaces, it is difficult to get the sub-surfaces which are statistically the same as the full aperture. For Figure 3.36 (a), the choice of sub-apertures leads to varying PSD calculations.

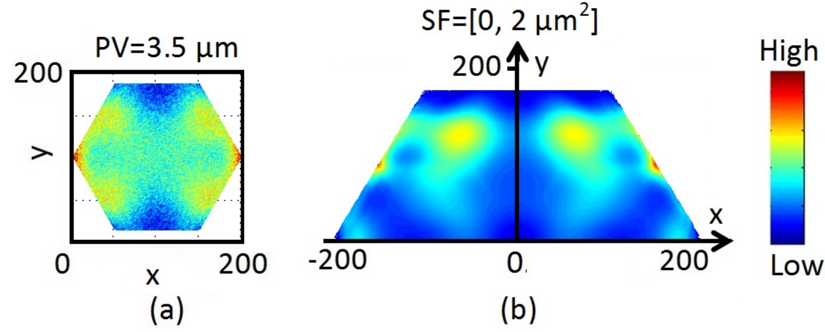


Figure 3.36 Analysis of simulation data.(a) Input map. (b) Area SF in 2 quadrants.

3.4 Root Mean Square (RMS) Gradient and Structure Function

As the SF is used to analyze the spatial frequency components of a surface, it is very important to connect the SF to the RMS gradient. The RMS gradient is the RMS value of the surface slope within the sampling area:

$$\begin{aligned}(\nabla z)_{RMS} &= \sqrt{\left\langle \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 \right\rangle_A} \\ &= \sqrt{\frac{1}{(m-1)(n-1)} \sum_{i=2}^{m}\sum_{j=2}^{n} \left\{ \left(\frac{z(i,j)-z(i-1,j)}{\Delta x}\right)^2 + \left(\frac{z(i,j)-z(i,j-1)}{\Delta y}\right)^2 \right\}},\end{aligned} \quad (3.68)$$

where Δx and Δy are the sampling intervals of the surface measurements in x and y dirctions respectively, m and n are the total number of sampling points in x and y dirctions respectively, and $z(i, j)$ are the surface height data.

Similar to the two-quadrant area SF, Eq. (3.68) only calculates the slope in the $(+x,+y)$ and $(-x,-y)$ directions. When calculated in the $(-x,+y)$ and $(+x,-y)$ directions, the RMS gradient is given as:

$$(\nabla z)_{RMS} = \sqrt{\left\langle \left(\frac{\partial z}{-\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 \right\rangle_A} = \sqrt{\left\langle \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 \right\rangle_A}$$
$$= \sqrt{\frac{1}{(m-1)(n-1)} \sum_{i=1}^{m-1} \sum_{j=2}^{n} \left\{ \left(\frac{z(i,j) - z(i+1,j)}{\Delta x}\right)^2 + \left(\frac{z(i,j) - z(i,j-1)}{\Delta y}\right)^2 \right\}}. \quad (3.69)$$

Apparently, the values of these two RMS gradients are the same.

In order to connect to the RMS gradient, the SF is expressed as [78]:

$$SF(r) = \frac{1}{A} \int \frac{1}{C} \int \left[z(P_1) - z(P_2) \right]^2 dP_1 dP_2, \quad (3.70)$$

where $z(P_1)$ is the surface height in the position P, A is the area of the surface, C is a circle with a radius $|P_1 - P_2| = r$ and centered at P_1, dP_1 represents the averaging over the area A, and dP_1 indicates the averaging along the circle C.

For a small separation r, Eq. (3.70) can be transformed to

$$SF(r \to 0) = r^2 \left\{ \frac{1}{A} \int \frac{1}{C} \int 2 \left[\frac{z(P_1) - z(P_2)}{r} \right]^2 dP_1 dP_2 \right\} = r^2 \left\{ \frac{1}{A} \int \frac{1}{C} \int \left(\frac{\partial z}{\partial r} \right)^2 dP_1 dP_2 \right\}$$
$$= \frac{r^2}{2} \left\langle \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 \right\rangle_A = \frac{r^2}{2} (\nabla z)^2_{RMS}. \quad (3.71)$$

The SF can also be described in the square-root way:

$$S(r) = \sqrt{SF(r)}. \quad (3.72)$$

Therefore, the SF slope at the origin [78] can be obtained with the RMS gradient of the surface:

$$S'(0) = \frac{(\nabla z)_{RMS}}{\sqrt{2}}. \quad (3.73)$$

3.5 Summary

This chapter investigates the relationship between the SF and other surface characterization techniques: Zernike polynomials, ACF, PSD, and RMS gradient.

According to the orthogonality of the Zernike polynomials, the total linear SF is the sum of the individual linear SFs of each of the Zernike polynomials with different azimuthal frequencies. However, this theorem does not apply to the area SF, because each area SF $S(r,\varphi)$ corresponds to only one direction φ, which means the orthogonality cannot be used.

The relationship between linear/area SF and linear/area ACF has been built for stationary surface, respectively. The SF contains similar information as ACF, but it provides better visualization of the surface characteristics than ACF, which is strongly dependent on the mean plane and is not intuitive.

Several important issues of the calculation of linear and area PSD have been discussed. Based on the relationship between the ACF and PSD, the SF was connected to the PSD in both linear and area domain.

The comparison between the SF and the PSD has been made.

Since the linear PSD is normally calculated along profiles, it uses only a small fraction of the surface data, which may not be the representative of the surface. While the linear SF is estimated over the whole surface, it reflects the true characteristics. Moreover, the linear PSD is sensitive to details of calculation methods, such as windowing, zero padding, and low-order terms removal, while the calculation of linear SF is definite and standardized.

Similar to linear PSD, the area PSD always needs additional mathematical processing, while the area SF is computationally “correct” for any arbitrary aperture shape without extra processing. There is no standardization of the mathematical treatment of PSDs for optical surfaces; different metrologists may compute different PSDs from the same starting data.

After connecting the SF to the RMS gradient, the SF slope at the origin has been evaluated.

CHAPTER 4 : COMBINATION OF AREA STRUCTURE FUNCTIONS FROM MULTIPLE INSTRUMENTS

Modern optical surfaces usually contain features over a wide range of spatial frequencies. In order to obtain a thorough analysis, multiple instruments with different upper and lower spatial frequency bandwidth limits have to be used. For the past 25 years, scientists (J. M. Bennett, P. Z. Takacs, E. L. Church, etc) have made comparison and combination of measurements from various types of instruments [34,97].

Currently, PSD is used to characterize spatial frequency component. The different PSDs of measurements from multiple instruments can be plotted on the same scale. However, the SF is not an orthogonal representation so combination of data from different instruments requires more than simply plotting on the same scale. This means, for example, a surface containing form error will have a non-zero SF for small separations [98]. Hence, higher magnification sub-aperture data must contain low order form for a correct ASF analysis.

When looking at the combination of the SF from multiple instruments, within the overlapped region in the response (spatial terms) of two instruments, there is often a discrepancy between instruments greater than the noise in the measurements. One cause for this discrepancy in optical instruments is that the response may be non-linear, showing a decrease in fidelity of the measurement of height as the spatial wavelength decreases. For smooth surfaces, this response can be characterized by the instrument transfer function (ITF) [99, 100, 101, 102, 103, 104, 105].

This chapter investigates the combination of the area SF from multiple instruments and analyzes the discrepancy with the ITF. Section 4.1 defines the ITF. Section 4.2 introduces the challenge of combination of SFs from multiple instruments: the area SF of the sub-aperture depends on the measured position. Section 4.3 discusses the convergence of sampling method based on simulated surfaces (both stationary and non-stationary surfaces). Section 4.4 describes the combination of area SFs from a Fizeau interferometer and a coherence scanning interferometer (CSI). Section 4.4.1 is based on a polished silicon (stationary surface), while section 4.4.2 is based on a diamond turned aluminum (non-stationary surface).

4.1 Instrument Transfer Function (ITF)

Basically speaking, the ITF is the ratio of the measured value to the true value. It represents the system response as a function of the input signal's frequency component. In surface metrology, a typical ITF is the optical transfer function (OTF), which describes the optical system's ability to reproduce images in terms of spatial frequency (reciprocal of lateral feature size).The modulus of the OTF is the well-known modulation transfer function (MTF) [99].

A convenient way to measure the ITF is to image a sharp reflectivity step [106, 107]. After setting imaging focus for the step and measuring it in a horizontal or vertical direction, the one-dimensional ITF can be estimated by taking the square root of the ratio of the PSD from the measured step to the PSD from the ideal step [101]:

$$ITF(f_{x,y}) = \sqrt{\frac{PSD(f_{x,y})_{measured}}{PSD(f_{x,y})_{idal}}}. \qquad (4.1)$$

For example, Figure 4.1 shows the theoretical and experimental ITF curves for four objectives used in the interferometric microscope. The horizontal dashed line at 70% of the ITF indicates the highest spatial frequency accepted by each objective [100].

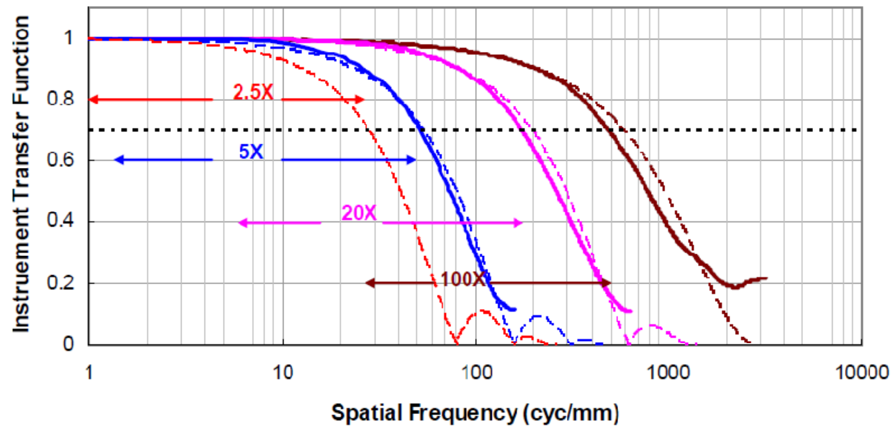


Figure 4.1 The theoretical (dashed lines) and experimental (solid lines) ITF for four objectives used in the interferometric microscope [100].

In general, the ITF shows the decreasing response as surface frequencies approach Nyquist, and the initial roll-off might begin at ~0.1 Nyquist with a well-designed instrument. If the surface contains high spatial frequency component, an instrument with higher lateral resolution is needed.

4.2 Challenge of Combination of SFs from Multiple Instruments

The main goal of this chapter is to demonstrate a methodology for combination of data from instruments with substantially different lateral dynamic ranges. Particularly, two instruments are discussed, a Zygo Verifire AT* Fizeau interferometer in Figure 4.2 (for full-aperture measurement) and a Zygo ZeGage* environmentally tolerant coherence scanning interferometer (CSI) in Figure 4.3 (for sub-aperture measurement).

---

*Commercial products identified to fully specify the experimental procedure

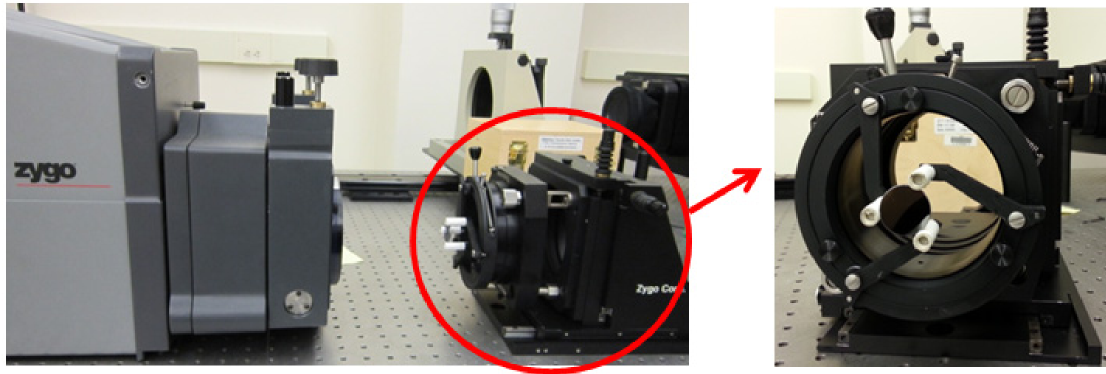


Figure 4.2 Measurement with Zygo Verifire AT Fizeau interferometer.

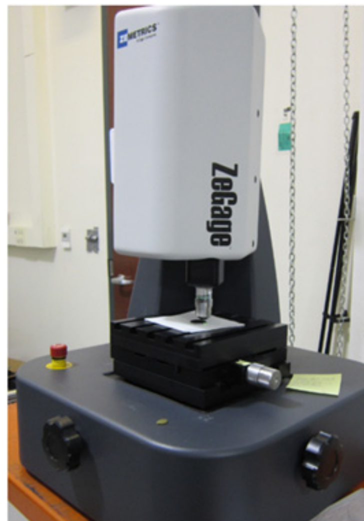


Figure 4.3 Measurement with Zygo ZeGage CSI.

For stationary surface, each sub-aperture has the same statistical characteristics, it is relatively easy to combine the sub-aperture measurement (relative high spatial frequency) and the full-aperture measurement (relative low spatial frequency). Figure 4.4 (b) is the full-aperture of a silicon surface, and Figure 4.4 (a) and (c) are two sub-apertures from different positions of the surface.

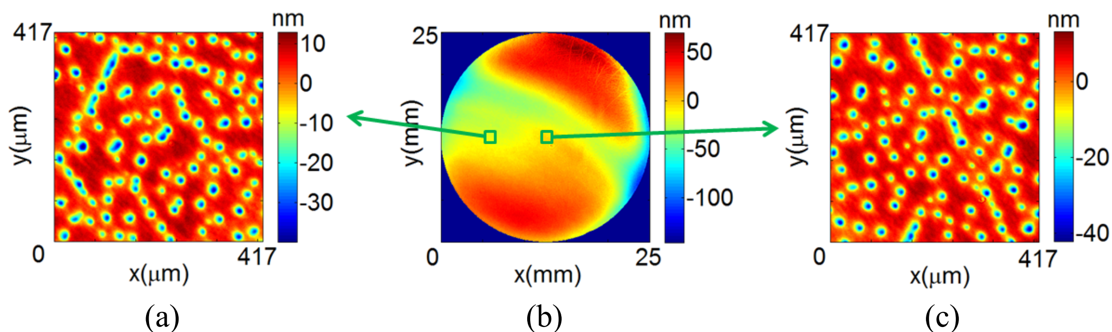


Figure 4.4 Silicon surface measured with different instruments. (a) A sub-aperture measurement with CSI at 20x. (b) Full-aperture measurement with Fizeau interferometer. (c) The other sub-aperture measurement with CSI at 20x.

Figure 4.5 depicts the RMS, skewness, and kurtosis of the two sub-apertures in Figure 4.4 (a) and (c). The RMS of the two sub-apertures are almost the same, and the skewness and kurtosis are very close between these two sub-apertures.

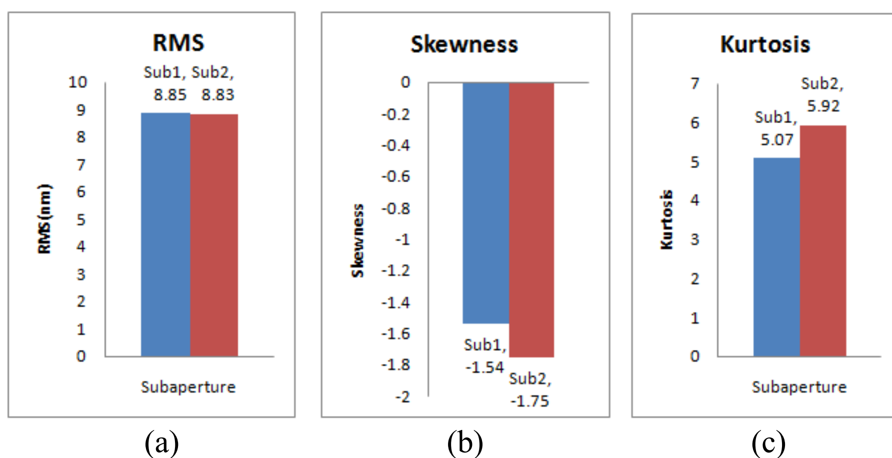


Figure 4.5 Some parameters of the two sub-apertures. (a) RMS of the two sub-apertures. (b) Skewness of the two sub-apertures. (c) Kurtosis of the two sub-apertures.

Figure 4.6 shows the area SFs of the two sub-apertures in Figure 4.4 (a) and (c). It is clear that the two area SFs contains similar spatial content.

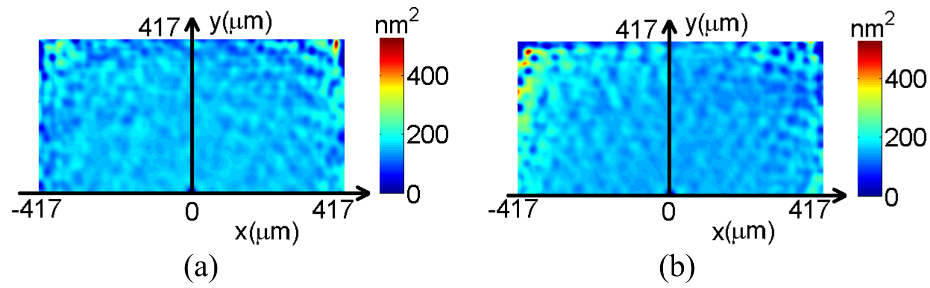


Figure 4.6 Area SFs of the two sub-apertures. (a) Area SF of the sub-aperture in Figure 4.4 (a). (b) Area SF of the sub-aperture in Figure 4.4 (c).

For non-stationary surface, the statistical characteristics of any sub-aperture is position dependent. For example, Figure 4.7 (b) is the full-aperture of a diamond turned aluminum surface, and Figure 4.7 (a) and (c) are two sub-apertures from different positions of the surface. As the turning marks dominate surface characteristic, the two sub-apertures are totally different.

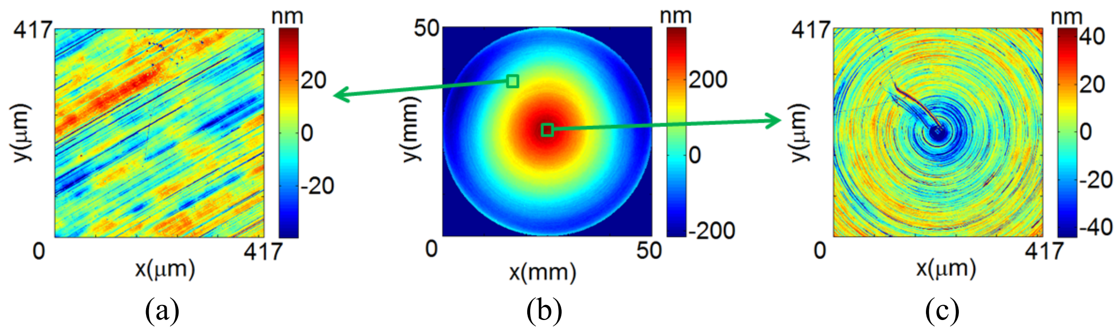


Figure 4.7 Diamond turned aluminum measured with different instruments. (a) A sub-aperture measurement with CSI at 20x. (b) Full-aperture measurement with Fizeau interferometer. (c) The other sub-aperture measurement with CSI at 20x.

Figure 4.8 shows the RMS, skewness, and kurtosis of the two sub-apertures in Figure 4.7 (a) and (c). The RMS and kurtosis of the two sub-apertures are similar. However, the difference between the two skewness is significant. That is because the skewness quantifies the symmetry of the surface topography and the height distribution of

the two sub-apertures are totally different. In brief, the two sub-apertures do not contain the same statistical characteristic.

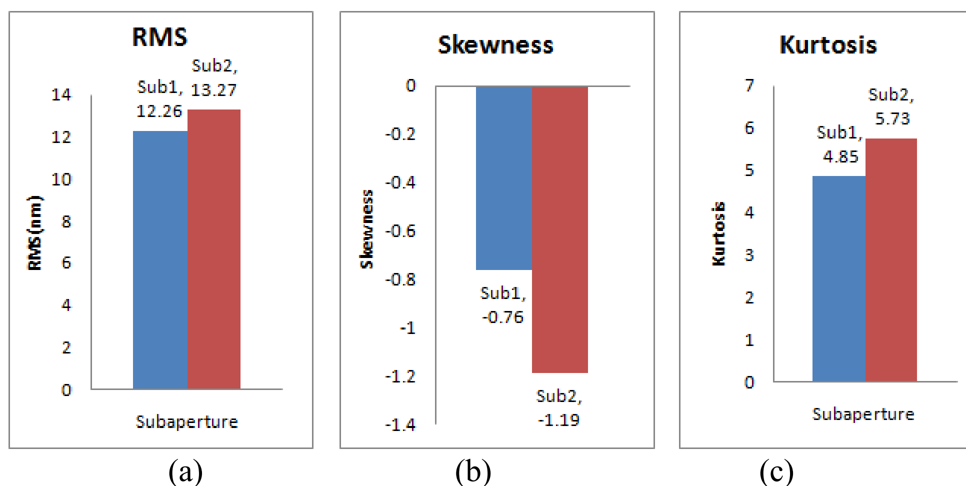


Figure 4.8 Some parameters of the two sub-apertures. (a) RMS of the two sub-apertures. (b) Skewness of the two sub-apertures. (c) Kurtosis of the two sub-apertures.

As Figure 4.9 shows, the area SFs of the two sub-apertures are totally different. Thus, for non-stationary surfaces, the area SF of the sub-aperture depends on the measured position.

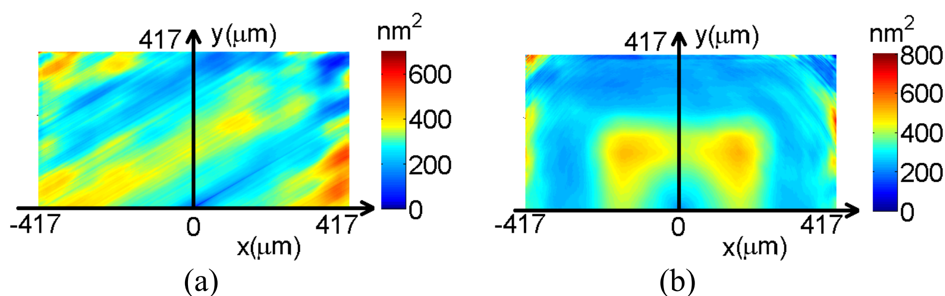


Figure 4.9 Area SFs of the two sub-apertures. (a) Area SF of the sub-aperture in Figure 4.7 (a). (b) Area SF of the sub-aperture in Figure 4.7 (c).

The 'brute force' approach to calculating the area SF for small separations (high spatial frequency) would be to measure the entire part surface area with the high lateral-resolution instrument, stitch together the sub-apertures, and then compute the area SF. This may require thousands of measurements and is experimentally and

computationally impractical. More realistically, the area SF can be estimated from the average of the area SF for a number of sampled sub-apertures. The details are discussed in the following section.

4.3 Convergence for Simulated Surfaces

Based on the simulation of a non-stationary surface and a stationary surface, this section demonstrates that the averaged area SF for a number of sampled sub-apertures converges to the analytically correct area SF for the high frequency content.

4.3.1 Non-stationary Surface

Figure 4.10 (a) shows a simulated sinusoidal surface (1000 μm×1000 μm) with a radial wavelength of 20 μm, and Figure 4.10 (b) shows its area SF. When zooming in the area SF near the zero separation, the high spatial frequency component is shown in Figure 4.10 (d).

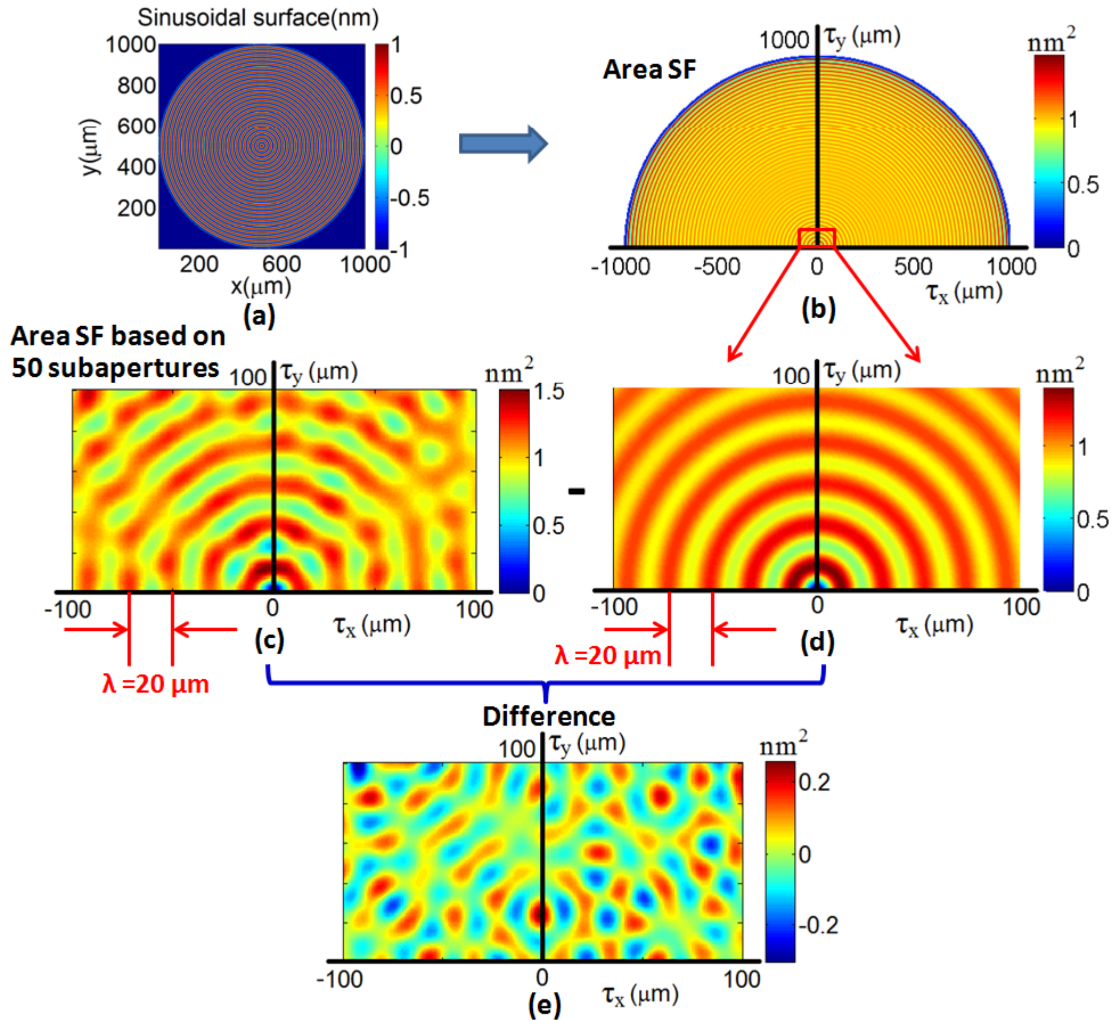


Figure 4.10 Sampling method. (a) Simulated sinusoidal surface. (b) Area SF of the simulated surface. (c) Average of the SFs of the sub-apertures. (d) Zoom in the area SF. (e) Difference between (c) and (d).

Fifty sub-apertures were picked randomly to simulate measurements taken with high magnification (100 μm×100 μm areas). Then, the area SFs were calculated for each sub-aperture and then averaged (shown in Figure 4.10 (c)). After comparing Figure 4.10 (c) and (d), the difference is shown in Figure 4.10 (e). It is clear that the averaged SF is an approximation of the exact SF in the high spatial frequency region shown in Figure 4.10 (d), they both represent the most important characteristic of the surface (λ = 20 μm). Thus, averaging the area SFs from sub-aperture measurements using instruments with high

spatial frequency resolution may represent the area SF in the high spatial frequency region even for a non-stationary surface.

In practice, the metrologist must estimate the number of sub-aperture samples needed to reach a given uncertainty. This convergence can be evaluated by looking at the difference between averaged sub-aperture SF and full-aperture SF as a function of N (number of sub-aperture samples). For example, Figure 4.10 (e) is the difference with N=50.

For the simulated radial sinusoidal surface, Figure 4.11 shows the area SF convergence for the sub-apertures with two sizes (100 μm×100 μm and 50 μm×50 μm), respectively.

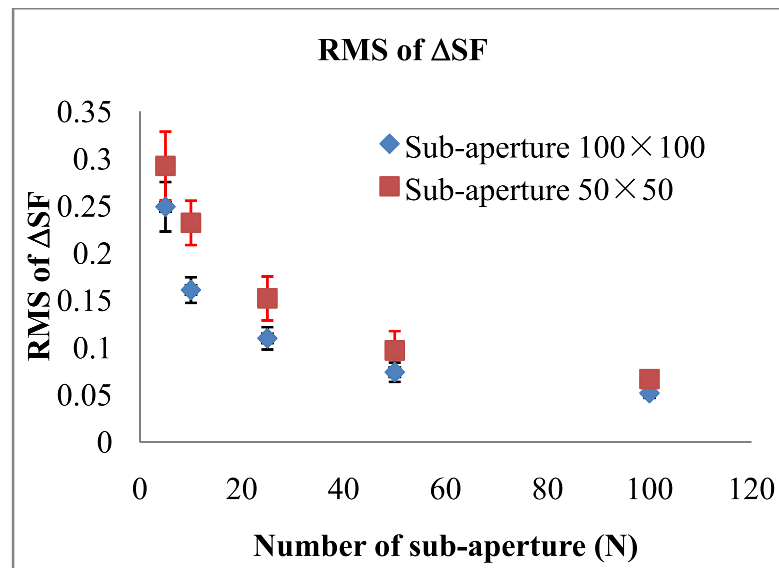


Figure 4.11 Area SF convergence for simulated radial sinusoidal surface.

For different values of the number of sub-apertures to be averaged (N), 10 independent groups of N sub-apertures were randomly selected and their area SFs were calculated and then averaged. The RMS of the difference between averaged sub-aperture area SF and

full-aperture area SF are plotted in Figure 4.11 as a function of *N*, showing the mean value for 10 groups and the standard deviation.

As expected, the RMS of ΔSF converges toward zero as N increases. For each sampling number of sub-aperture, the RMS of ΔSF with size 100 μm×100 μm is less than that with size 50 μm×50 μm, because larger area contains more information during the calculation. In this case, the surface contains periodic data with only one spatial frequency. When the size of the sub-aperture is reduced from 100 μm×100 μm to 50 μm ×50 μm, the number of pixels is reduced by a factor of 4 and the number of cycles is reduced from 5 to 2.5 per sub-aperture. As noted in Chapter 2, the periodic maximum SF values are not exactly separated by the spatial wavelength because of finite sampling and averaging effects, especially for the values at the edge. Therefore, the SFs of sub-apertures containing fewer pixels and cycles have more noise and errors. Thus, the convergence to zero is faster when the area of the sub-aperture is increased.

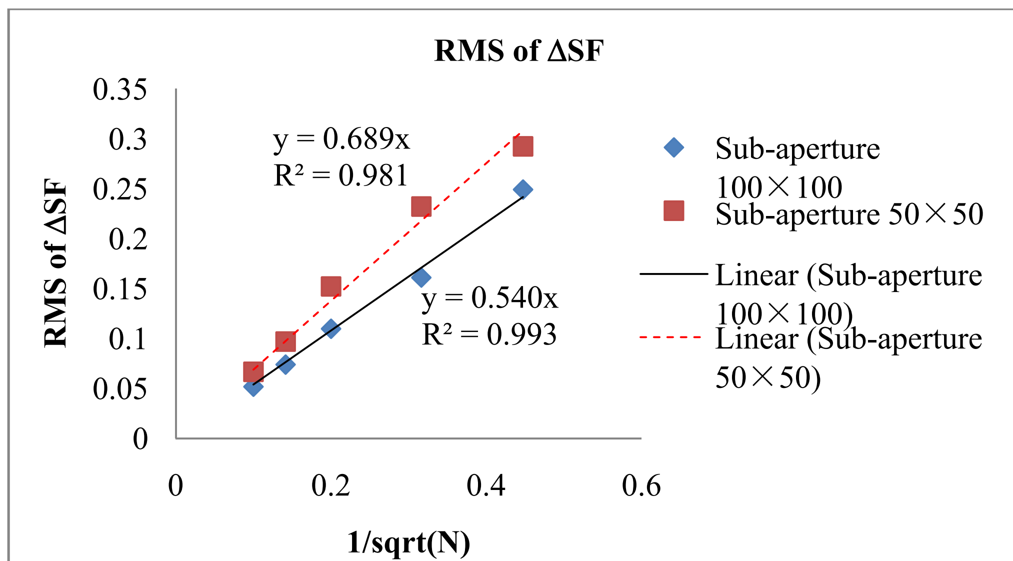


Figure 4.12 Area SF convergence indicates a dependence consistent with $(N)^{-1/2}$ for simulated sinusoidal surface.

Replotting Figure 4.11, the convergence shows a dependence consistent with $(N)^{-1/2}$ in Figure 4.12.

4.3.2 Stationary Surface

The convergence of stationary surface was also investigated.

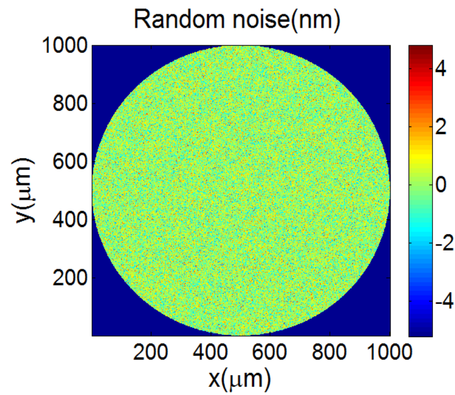


Figure 4.13 Simulated random noise.

For the simulated random noise (Figure 4.13), Figure 4.14 and Figure 4.15 show the area SF convergence for the sub-apertures with two sizes (100 μm×100 μm and 50 μm× 50 μm), respectively.

As with the non-stationary surface, the RMS of ΔSF converges to zero for increasing *N* and converges faster for larger sub-apertures.

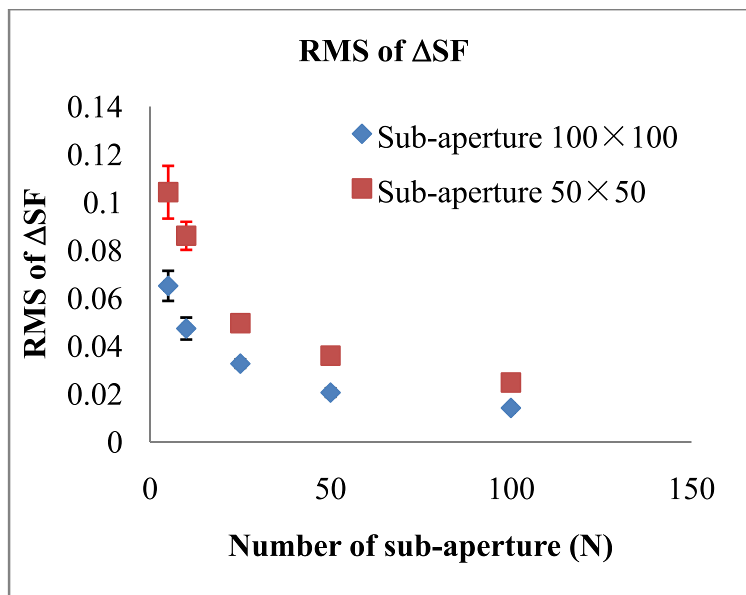


Figure 4.14 Area SF convergence for simulated random noise.

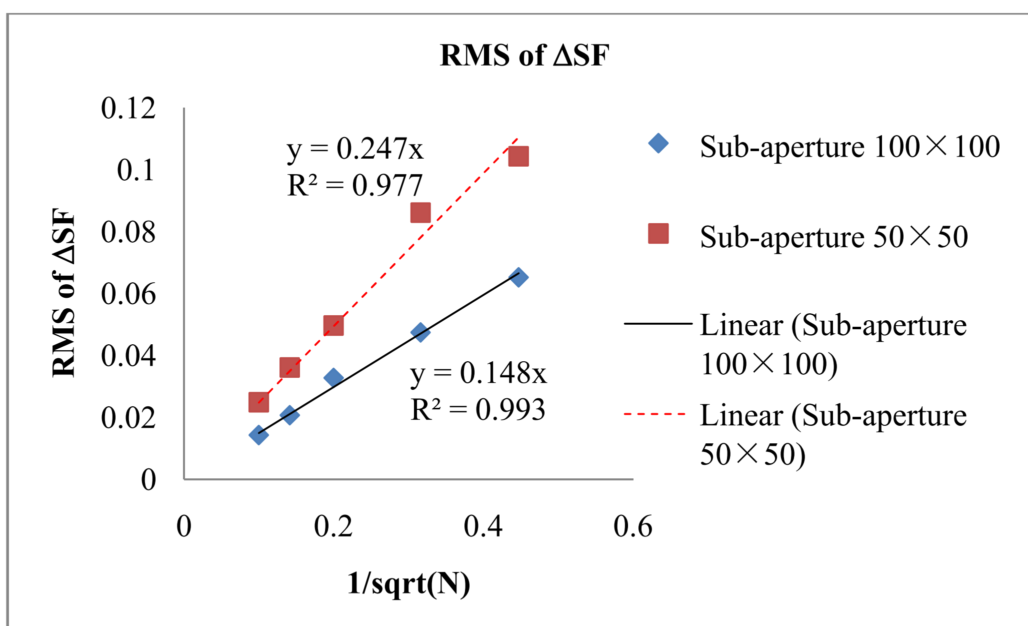


Figure 4.15 Area SF convergence indicates a dependence consistent with $(N)^{-1/2}$ for simulated random noise.

In general, the convergence for the stationary surface is faster than the non-stationary surface, because each sampled sub-aperture of the non-stationary surface may contain different statistical characteristics.

4.3.3 Tilt Removal

One important characteristic of an area SF is that it is not an orthogonal representation of the spatial content of the surface in terms of separation. This means, for example, a surface containing form error will have a non-zero SF for small separations. Hence, higher magnification sub-aperture data must contain low order form for a correct area SF analysis. Take the CSI measurement as an example, tilt is the only low order form contribution that is unknown for sub-aperture measurements. This is because the apparent tilt is assumed to be an alignment error and is usually removed [108]. The correct area SF can be estimated by adding the correct tilt to each sub-aperture measurement, perhaps estimated from the large-area measurement and knowledge of the global coordinates of each. For a purely quadratic parent surface, the sub-apertures will appear to have variable tilt as a function of position and the same power. Note that the ratio of the power in the parent and in each sub-aperture is proportional to the square of the aperture sizes and is independent of position.

For example, Figure 4.16 (a) shows a simulated surface map that consists of power, coma and radial sine wave. Figure 4.16 (b) and (c) show its area SF. Figure 4.16 (d) shows the randomly picked sub-apertures (including the correct tilt), and Figure 4.16 (e) shows the averaged SF. We see that Figure 4.16 (e) is an estimate of Figure 4.16 (c).

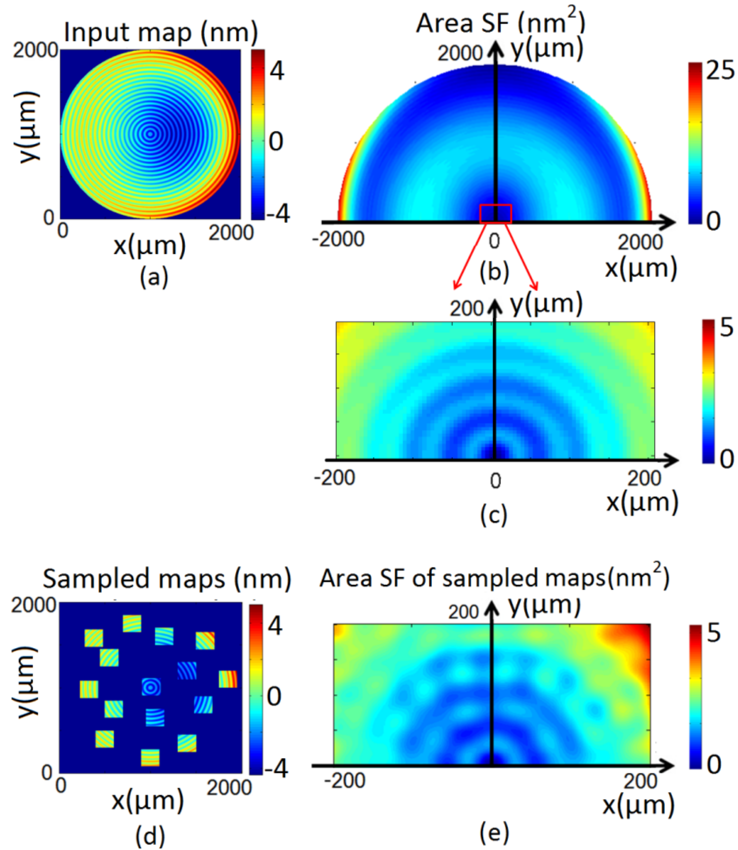


Figure 4.16 Tilt analysis of a simulated surface. (a) Simulated surface map with power, coma and a sine wave. (b) Area SF of the simulated surface.(c) Zoom in on the area SF in the high-spatial frequency region. (d) Randomly picked sub-apertures with tilt. (e) Average of the SFs of the sub-apertures.

However, if the tilt of the randomly picked sub-apertures is removed, the averaged SF would be similar to that shown in Figure 4.10 (d). Notice that the wavelength of the sine wave is still identifiable.

4.4 Combination of Area Structure Functions

Following the simulation, two real precision surfaces (a polished silicon flat and a diamond-turned aluminum flat) were analyzed.

4.4.1 Polished Silicon Flat

4.4.1.1 Experimental Method

Form error of the polished silicon flat was measured using the Zygo Verifire AT Fizeau interferometer with a ring source. The system was zoomed so the part fills the 1000×1000 pixel detector array. Height data are exported for off-line computation.

Three fiducials on the edge of the part can be identified in the Fizeau data and in measurements made using a Zygo ZeGage environmentally tolerant coherence scanning interferometer (CSI) [109] with a 20x objective and a $(417\ \mu m)^2$ field of view. Position and rotation of the fiducials allows the coordinate system of the Fizeau data to be connected to the coordinate system defined by the stages of the CSI. This is necessary because tilt in individual sub-aperture CSI measurements is alignment sensitive and not uniquely related to the form of the optic; so tilt in the raw CSI data must be replaced by the correct tilt associated with the specific position in the full aperture. The CSI measurements capture quadratic and higher order components of form; hence only tilt needs to be corrected.

4.4.1.2 Uncertainty Associated with Tilt Correction

Figure 4.17 (a) is the polished Si flat measured with the Fizeau interferometer. The small square represents a sub-aperture CSI measurement at 20x. In order to estimate the tilt of the sub-aperture, a fitted surface with 36 Zernikes was generated (Figure 4.17 (b)). Figure 4.17 (c) is the sub-aperture measured with the CSI. Figure 4.17 (d) shows the area SF for the Fizeau data. The figure error is dominated by a comatic term (cubic, approximately symmetric about the *y*-axis); the area SF shows peaks at separations of

approximately half the aperture size, as well as peaks near full aperture separation resulting from the high and low edges at 3 and 9 o'clock.

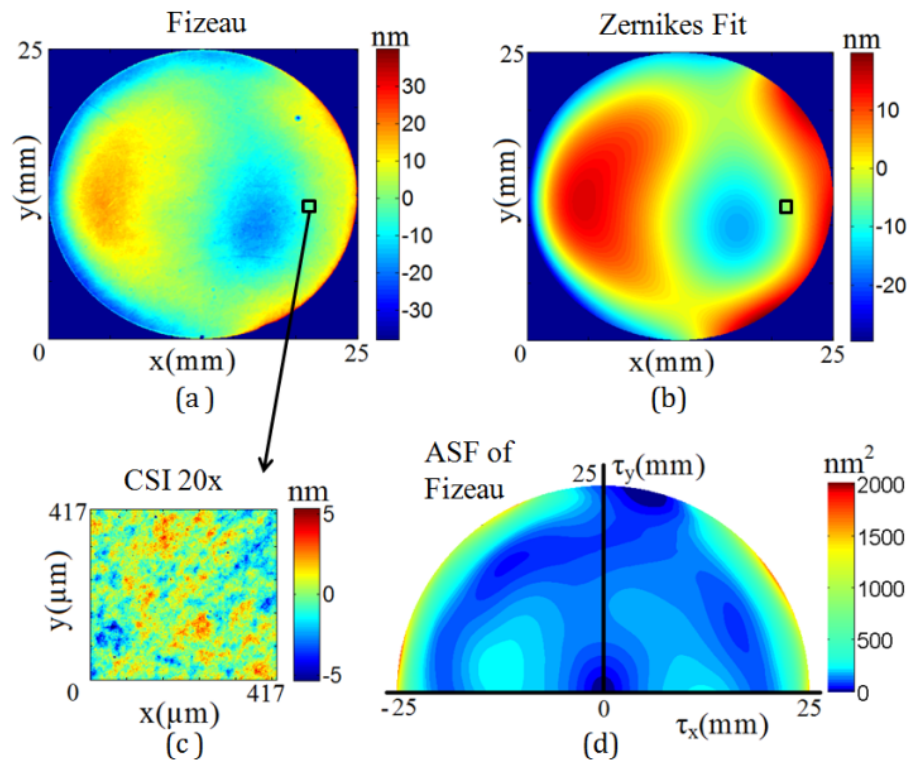


Figure 4.17 Polished silicon flat. (a) Polished silicon flat measured with Fizeau interferometer. (b) Fitted surface with 36 Zernikes. (c) A sub-aperture measured with the CSI at 20x. (d) Area SF of the Fizeau data.

With the fiducials, the coordinate system defined by the stages of the CSI can be connected to the coordinate system of the Fizeau measurement. During the CSI measurement, the stages are translated in x and y directions (recording the x and y scale). When comparing the recorded scale to the positions of the fiducials in the coordinate system of the Fizeau measurement, the position of the sub-aperture can be obtained.

Error in the estimate of the sub-aperture position leads to error in the tilt correction and therefore error in the area SF calculation. An analysis shows that the dominant uncertainties in positioning include the x and y positioning uncertainty in connecting the

two coordinate systems (±0.1mm), x and y calibration uncertainty of the linear axes of the CSI (±0.1mm), and the x and y axes squareness (±0.04mm). Other contributions such as the straightness, angular error motions of the stages and distortion in the Fizeau interferometer are negligible. The dominant uncertainties are uncorrelated and the expanded uncertainty (k=2) in sub-aperture position in the Fizeau data coordinates is:

$$U = 2\times\sqrt{(0.1)^2+(0.1)^2+(0.1)^2+(0.1)^2+(0.04)^2} = 0.4(mm) \quad (4.2)$$

Uncertainty in the tilt correction, for a given sub-aperture, arising from positioning uncertainty depends on the local curvature (ie change in slope). Figure 4.18 (a) is a curvature map for the fitted surface shown in Figure 4.17 (b). The uncertainty consequence for the area SF calculation was investigated by selecting sub-apertures on the surface at locations with different curvatures. For each nominal position, the area SF was computed using the nominal tilt and a modified area SF was computed using tilts for positions 0.4 mm away. The RMS of the difference between the area SF and the modified area SF was computed as this provides a reasonable metric to evaluate the effect. Figure 4.18 (b) summarizes the analysis and shows that, for this surface, the difference rises rapidly for curvature above 4×10^{-3} m^{-1}.

The short period structure on the silicon surface is relatively insensitive to position (see Figure 4.17 (c)) and only a small fraction of the surface has curvature exceeding 4×10^{-3} m^{-1}. Therefore, it is reasonable to use sub-apertures only from positions on the surface which have curvatures lower than this limit.

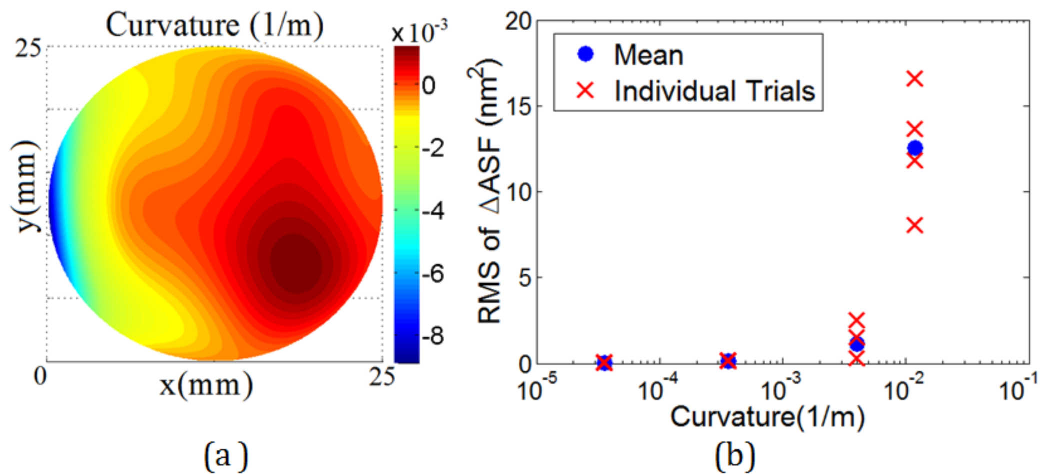


Figure 4.18 Curvature. (a) Curvature map of the fitted surface. (b) RMS of ΔSF vs. curvature.

4.4.1.3 Area SF Convergence for CSI Measurements (20X)

The purpose of combining data from two instruments is to provide robust estimates of the area SF from separations ranging from full aperture to the high spatial frequency limit of the CSI. As discussed in section 4.2, the area SF can be estimated from the average of the area SF for a number of sampled sub-apertures. Note that the average area SF is physically sensible, and has been shown in simulation to converge to the analytically correct area SF.

The convergence of the averaged area SF with the number of sub-aperture samples must be estimated to decide on the number of sub-aperture measurements needed.

In practice, because the lateral resolutions of sub-aperture measurement and full-aperture measurement are different, we cannot directly evaluate the convergence by looking at the difference between averaged sub-aperture SF and full-aperture SF.

Instead, the convergence can be evaluated by looking at the difference of independent samples of N averages from a larger population of measurements. This is demonstrated for the Si flat where 100 sub-apertures were measured with the CSI (20x objective). The

position of each sub-aperture was recorded for subsequent tilt correction. For different values of the number of sub-apertures to be averaged (N), independent groups of N measurements were randomly selected from the population and their tilt-corrected area SFs were calculated and then averaged. Finally, the RMS of the difference between pairs were plotted as a function of N, showing individual trials and the mean. For example, Figure 4.19 shows the flow chart of calculating RMS of ΔSF for N=25.

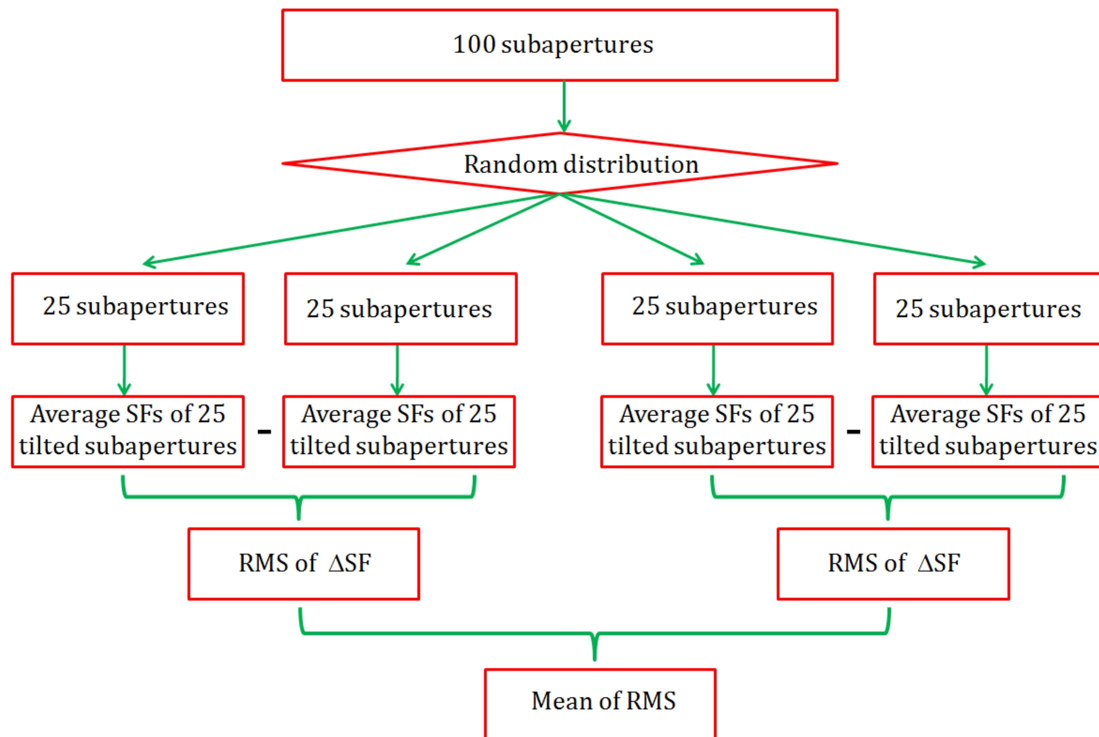


Figure 4.19 Flow chart of calculating RMS of ΔSF.

Figure 4.20 shows the RMS of the difference between pairs as a function of N, including individual trials and the mean. Notably, the convergence shows a dependence consistent with $(N)^{-1/2}$.

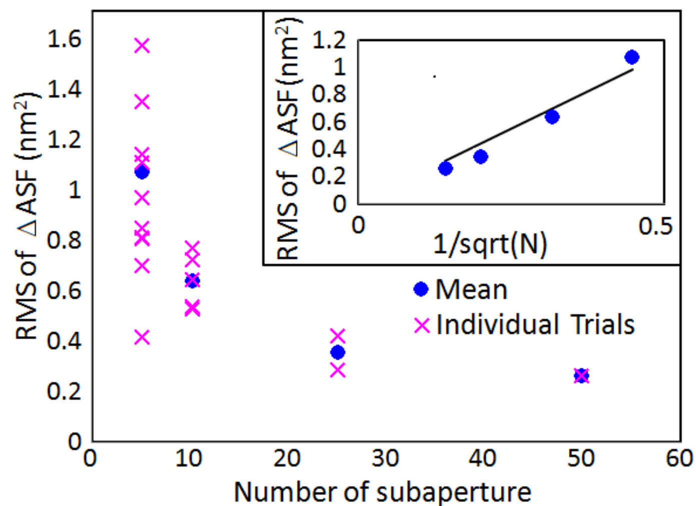


Figure 4.20 Area SF convergence for polished Si flat (CSI 20x objective).

In fact, the selection of N will depend on the surface, the size of the sub-aperture, and the specification. Following the default decision rule of ISO 14253-2 [110], the expanded uncertainty in the area SF must be subtracted from the specification for a vendor to prove conformance. Reducing the uncertainty increases the manufacturing headroom.

4.4.1.4 Combination of Area SFs from Two Instruments

Figure 4.21 (top left) shows the area SF for the silicon flat computed from the Fizeau data. Zooming in (top right) the area SF trends asymptotically to zero for zero separation, as expected. For real interferometers, the ITF shows the decreasing response as surface frequencies approach Nyquist, and the initial roll-off might begin at ~0.1 Nyquist with a well-designed instrument. Consequently, the ITF impacts the Fizeau data over much of the range shown in the top right. The averaged area SF from the CSI measurements is computed (mid-left in Figure 4.21) and substituted for the data over the corresponding separation range in the Fizeau data (Figure 4.21 bottom). Comparison of Figure 4.21 (bottom and top right) indicates the increased fidelity of the combined area SF. The

discontinuity between CSI and Fizeau data at the top corners of the CSI insert is a consequence of noise and low sampling and is not meaningful.

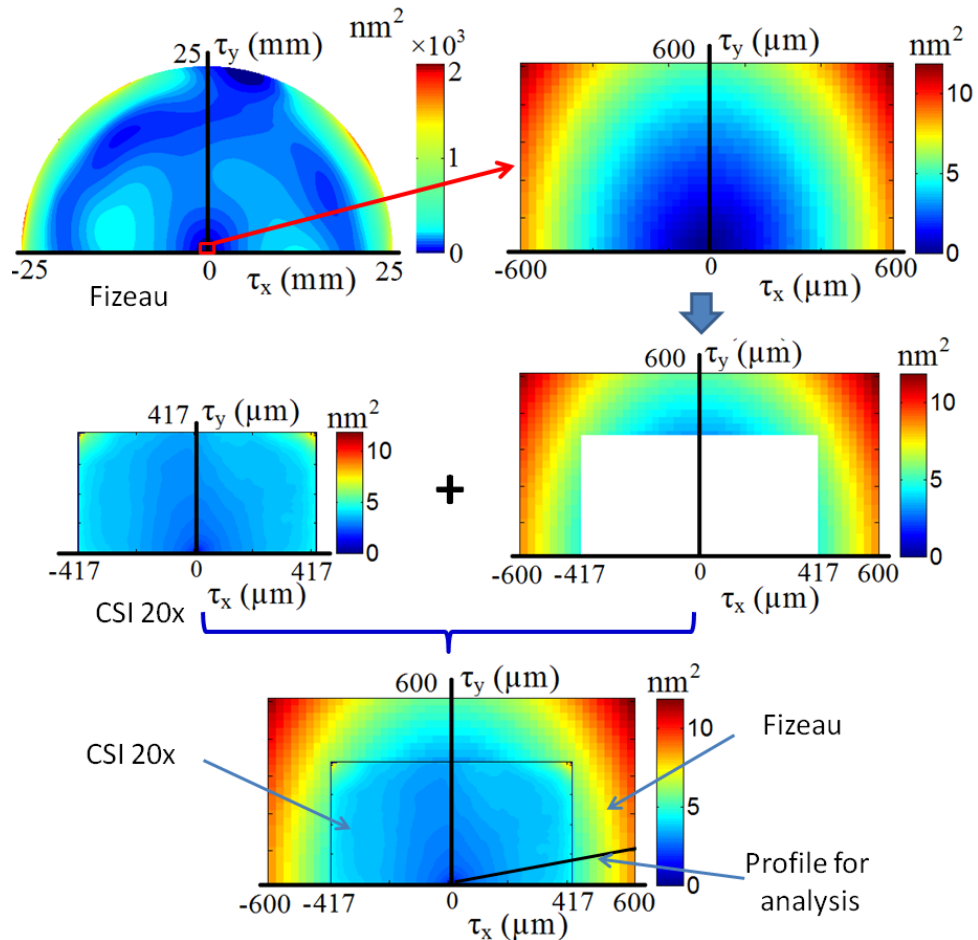


Figure 4.21 Combination of area SF based on the Si flat measured with the Fizeau interferometer and the CSI (20x).

4.4.1.5 Type A Uncertainty in Combined Area SF

Figure 4.22 extracts data for a specific profile (combination of *x* and *y* separations) for the combined area SF based on Fizeau and CSI data. The resolution of the Fizeau measurement is 25 μm/pixel while the CSI data is approximately 0.4 μm/pixel.

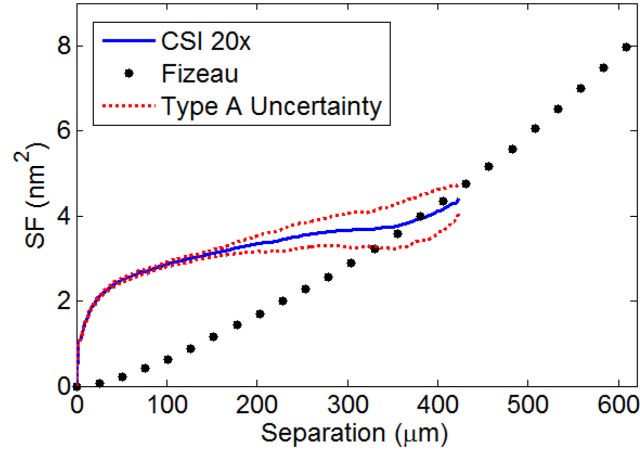


Figure 4.22 Variation for a profile from the combined area SF.

The uncertainty in the averaged area SF for the CSI data is dominated by the tilt-correction uncertainty and by the finite averaging uncertainty for larger values of tau. Both randomly vary among the N=50 samples in the population used for the average. Thus, a Type A uncertainty analysis can be used to give a good estimate of the combined standard uncertainty each τ as

$$u_{\tau} = \sqrt{\frac{\sum_{i=1}^{N}(SF(\tau)_i - SF(\tau)_{avg})^2}{N \times (N-1)}} \quad (4.3)$$

Applying a coverage factor k=2 for each data point, the Type A analysis is shown as a red dashed curve above and below the CSI data in Figure 4.22.

The results in Figure 4.21 and Figure 4.22 show that, at separations of approximately 400 µm, there is reasonable agreement between the SFs computed using the two instruments. This is consistent with the analysis with ITF: the resolution of the Fizeau measurement is 25 µm / pixel, giving close to 0.1 Nyquist at τ = 400 µm. In this case, the ITF is expected to be essentially unity [111]. At smaller values of τ (greater than 0.1

Nyquist), the Fizeau data falls below the CSI data as the ITF of the Fizeau approaches zero at the Nyquist limit.

4.4.1.6 Combination of Area SFs without Tilt Correction

The combination of area SFs without tilt correction is shown in Figure 4.23. The averaged area SF from the CSI measurements is computed (no tilt correction) and substituted for the data over the corresponding separation range in the Fizeau data.

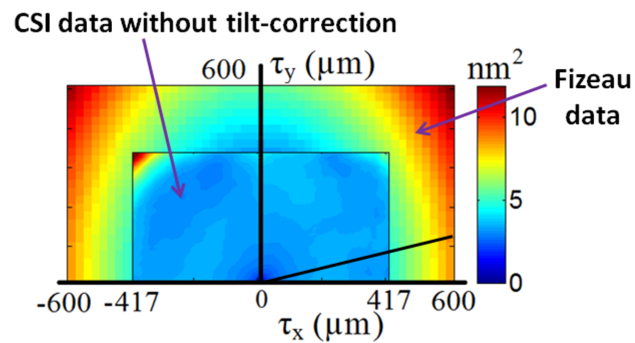


Figure 4.23 Combination of area SF based on the Si flat measured with Fizeau interferometer and CSI 20x (no tilt-correction).

Figure 4.24 extracts data for a specific profile (the same position as that in Figure 4.21) for the combined area SF from Fizeau and CSI measurements. It is obvious that the CSI data does not match as well with the Fizeau data as when tilt correction is made.

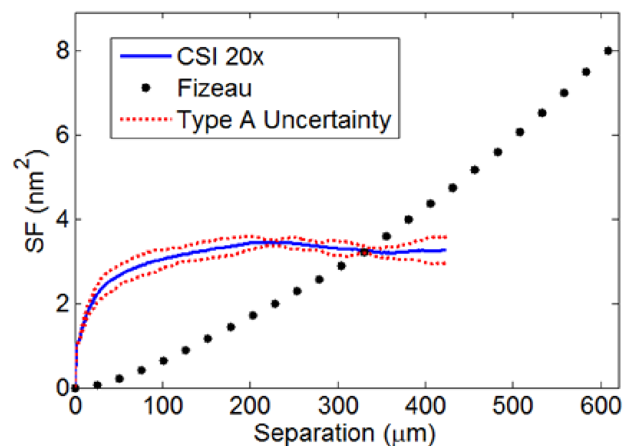


Figure 4.24 Variation for a profile from the combined area SF (no tilt-correction).

4.4.2 Diamond Turned Aluminum

4.4.2.1 SF Convergence for CSI Data (20X)

For nonstationary and highly anisotropic surfaces, it is possible to represent the area SF in the same way. A 50 mm diameter diamond turned aluminum flat was measured with both the Fizeau interferometer (Figure 4.25 (a)) and the CSI (20x objective).

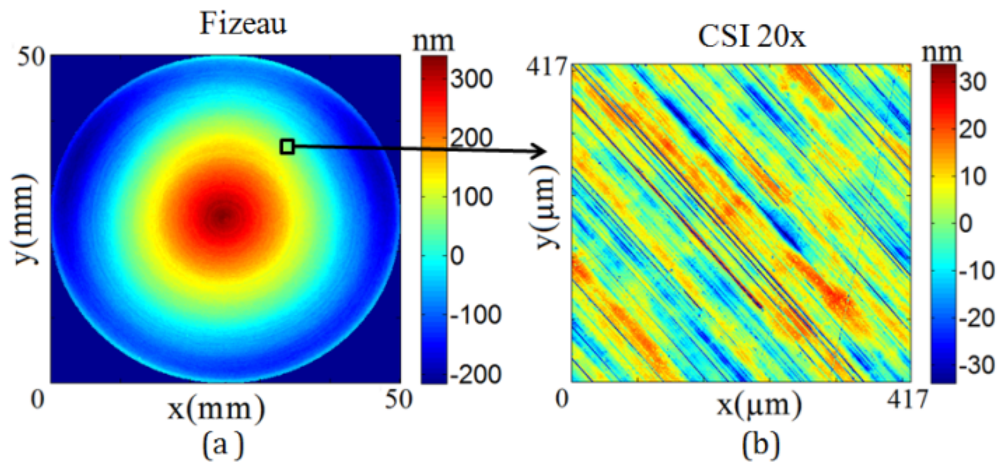


Figure 4.25 Diamond turned Al. (a) The whole surface measured with Fizeau interferometer. (b) A sub-aperture measured with the CSI 20x.

Following the same procedure as with the polished Si, 100 sub-apertures were measured with the CSI. The same trend of decreasing RMS of area SF difference with the number of sub-aperture is observed (Figure 4.26), although the amplitudes are much greater, as expected. The convergence to zero is slower due to the anisotropy in the high spatial content regime of the diamond turned surface.

By contrast, the polished silicon surface can be considered a stationary surface over the bandwidth of the CSI data, after removal of tilt. Thus the averaged area SF of the sub-apertures converges at a rate that appears to be a consequence of the low order form, which is faster than the diamond turned surface.

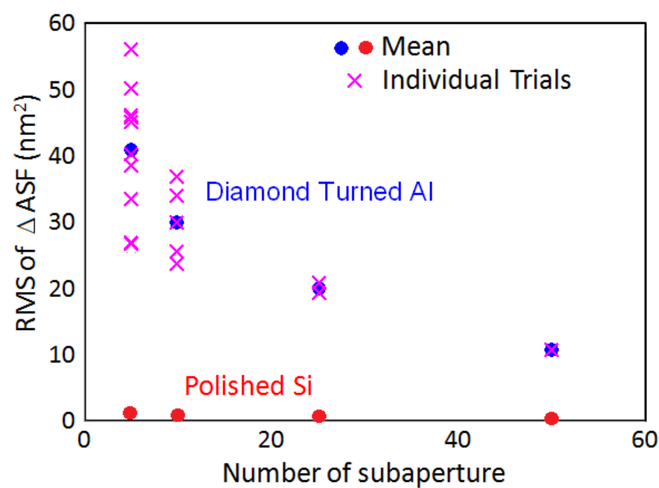


Figure 4.26 Area SF convergence for diamond turned Al and the comparison with polished Si flat (CSI 20x).

4.4.2.2 SF Combination from CSI (20X) and Fizeau Data

Similar to the polished Si, the area SF based on Fizeau data and the area SF based on CSI data are combined in the high spatial content region (small separation) in Figure 4.27 and Figure 4.28.

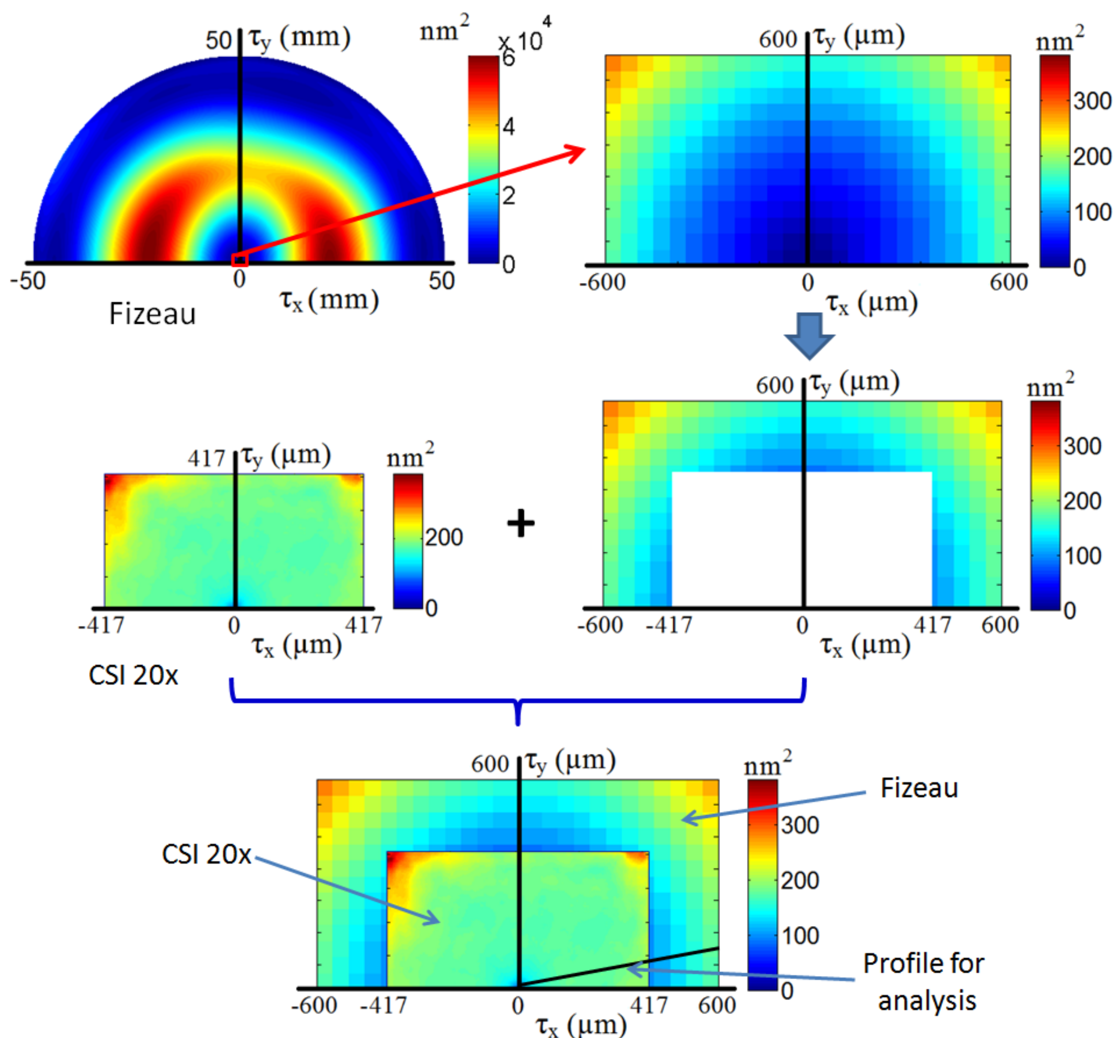


Figure 4.27 Combination of area SF based on the diamond turned Al measured with Fizeau interferometer and CSI (20x objective).

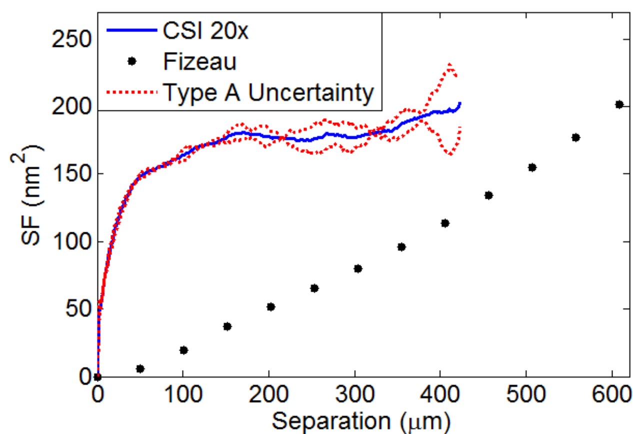


Figure 4.28 Variation for a profile from the combined area SF.

Figure 4.27 and Figure 4.28 show a significant difference between the Fizeau and CSI based area SFs for the diamond turned aluminum surface for τ values of order 0.4 mm. For the large τ values in the CSI data, the ITF of the instrument is essentially unity. For the Fizeau, in this case, each pixel is 50 μm square and so the Nyquist frequency is 1/100 μm^{-1}.

From Figure 4.28, we observe that the ratio of CSI to Fizeau SF is approximately $130nm^2$ / $200nm^2$ = 0.65 for τ= 400 μm, or 0.25 Nyquist. If the entire difference between the Fizeau and CSI SFs (nm^2) is attributed to the ITF of the Fizeau, this implies an ITF of $0.65^{1/2} \approx 0.8$ at 0.25 Nyquist. The ITF of the specific Fizeau has not been measured at this zoom; however, this estimated value is consistent with the published literature [99, 111] and reasonable according to the instrument manufacturer [112].

As we have discussed, the ITF shows the decreasing response as surface frequencies approach Nyquist, and the initial roll-off might begin at ~0.1 Nyquist with a well-designed instrument. For the Fizeau measurement in this case, 0.1 Nyquist is related to the separation of 1000 μm. Thus, if the separation range is extended to some value greater than 1000 μm, we can obtain a comprehensive evaluation of the area SF combination.

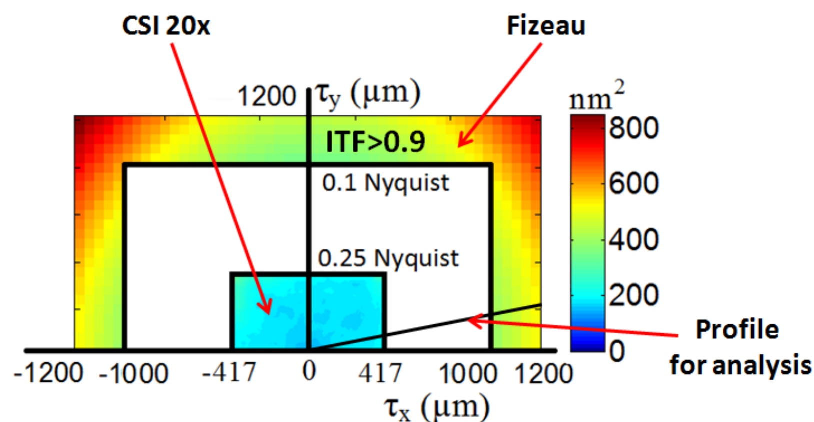


Figure 4.29 Combination of area SF based on the diamond turned Al measured with Fizeau interferometer and CSI (20x objective).

As Figure 4.29 shows, when the separation is greater than 1000 μm (less than 0.1 Nyquist frequency), the ITF of Fizeau measurement is greater than 0.9, which means the Fizeau data in this region is reasonable. When the separation is less than 417 μm (greater than 0.25 Nyquist frequency), the Fizeau data should be replaced by the CSI (20x) data. Between the separation of 417 μm and 1000 μm, there are no reasonable data.

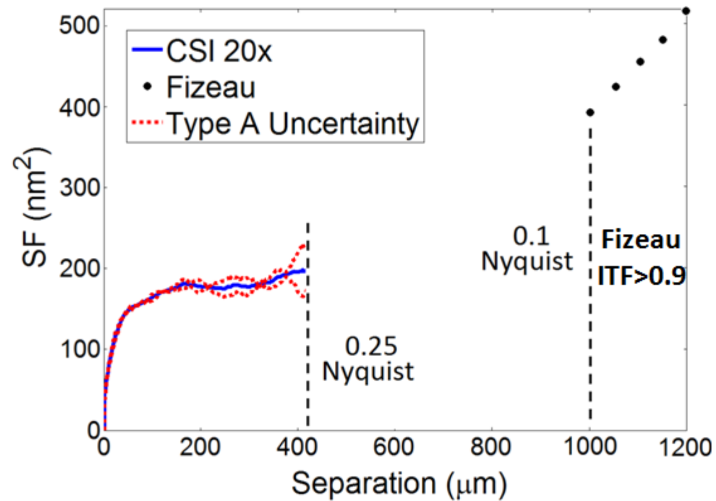


Figure 4.30 Variation for a profile from the combined area SF.

Figure 4.30 extracts data for a specific profile for the combined Fizeau and CSI (20x) area SF. If we want to cover the spatial frequency between 1/417 μm^{-1} and 1/1000 μm^{-1}, a lower-magnification CSI is needed.

4.4.2.3 SF Combination from CSI (5X and 20X) and Fizeau Data

In order to make a better combination with Fizeau SF, a 5x objective was selected for the CSI measurement. In this case, the resolution of the Fizeau measurement is 50 μm/pixel and the CSI (5x) measurement is approximately 1.6 μm/pixel.

Similar to the procedure in last section, the area SF based Fizeau data and the area SF based on CSI (5x) are combined in the high spatial content region (small separation) in Figure 4.31.

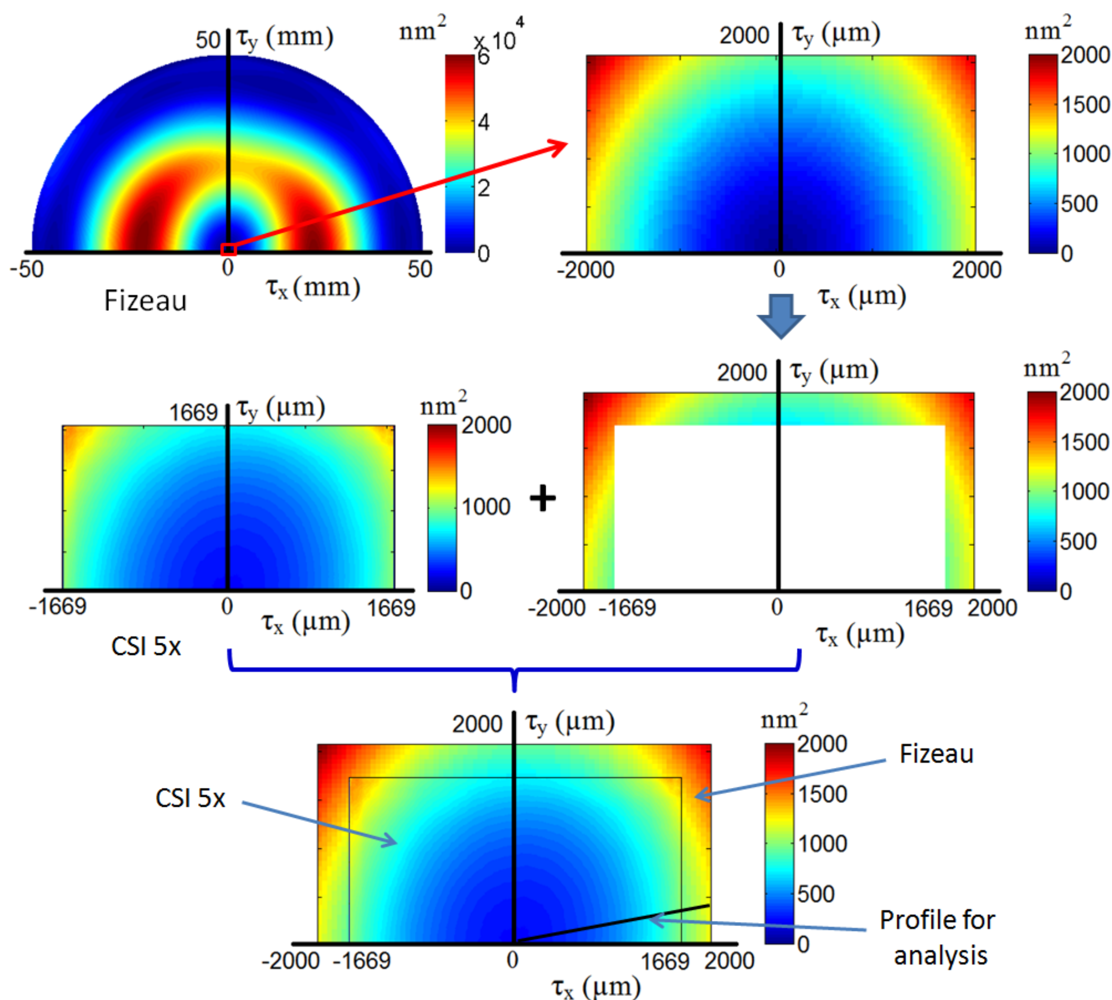


Figure 4.31 Combination of area SF based on the diamond turned Al measured with Fizeau interferometer and CSI (5x objective).

Figure 4.32 extracts data for a specific profile for the combined Fizeau and CSI (5x) SF. The result shows that the SFs computed using the two instruments have a reasonable agreement between the separations of approximately 1 mm (0.1 Nyquist) and 1.6 mm (0.06 Nyquist).

At smaller values of τ, the Fizeau data falls below the CSI data as the ITF of the Fizeau approaches zero at the Nyquist limit.

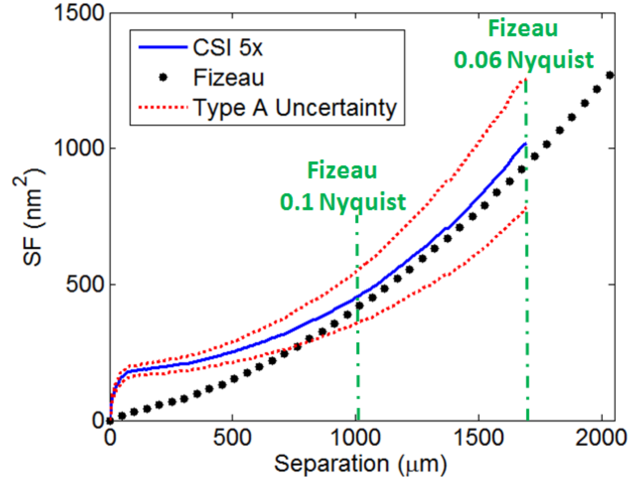


Figure 4.32 Variation for a profile from the combined area SF.

Furthermore, it is worthwhile to combine the SFs from CSI 20x, CSI 5x and Fizeau measurements. Figure 4.33 (b) shows the three SFs on the same scale.

As Figure 4.33 (a) shows, there is a discrepancy between the SFs from CSI 5x and CSI 20x, but the mean estimate SF for the 20x data falls with the Type A uncertainty (expanded at k=2). The uncertainties in the averaged area SFs for the CSI data are dominated by the tilt-correction uncertainty and by the finite averaging uncertainty for larger values of separation. Both randomly vary among the N=50 samples in the population used for the average.

Therefore, to reduce the discrepancy between the two SFs, it is necessary to improve the tilt-correction (i.e. reduce the error in the estimate of the sub-aperture position) and increase the number of sub-aperture measurements.

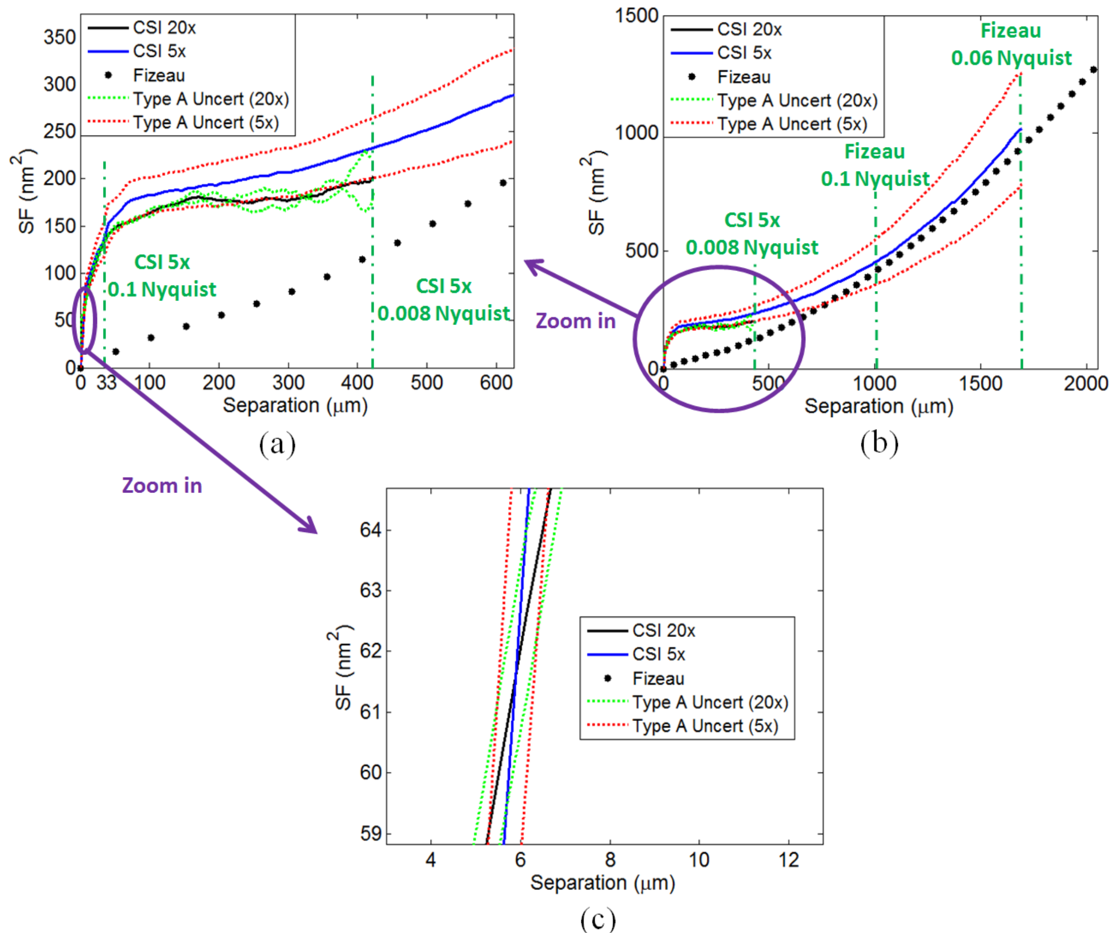


Figure 4.33 Combination of the SF for Fizeau data and CSI data (at 5x and 20x). (a) Zoom in SF in the high spatial frequency region. (b) Variation for a profile from the combined area SF (Fizeau, CSI 5x, and CSI 20x). (c) Intersection of the CSI SFs (5x and 20x).

When zooming the SF in the high spatial frequency region (Figure 4.33 (c)), it is clear that the SF of CSI 5x falls below the SF of CSI 20x as the separation is less than 6 µm. This is because the ITF of the CSI 5x approaches zero at the Nyquist limit.

Note also that Type B uncertainties have not been considered. In neither set of measurements has the reference surface of the objective been calibrated, nor distortion evaluated and corrected.

4.5 Summary

The spatial content of surfaces over large dynamic ranges for arbitrary aperture shapes can be represented by an area SF constructed from data obtained from more than one instrument. Proper combination of area SFs from different instruments requires accounting for the effect of low order form on small separation SF values. The effect of the ITF of the instruments used must also be considered.

A full uncertainty analysis on a combined area SF should include Type B uncertainties, such as the effect of an uncalibrated reference flat or retrace error, not considered here.

CHAPTER 5 : APPLICATION OF STRUCTURE FUNCTION

Besides the applications mentioned in Chapter 1, the structure function (SF) can also been used in other surface metrology applications. With the SF, this chapter explores two typical noise sources (electronic noise and air turbulence) for the phase shifting interferometry (PSI), and makes a quantitative estimation of a diamond turned surface after the compensation machining based on a dynamic measurement.

5.1 Investigation of Electronic Noise and Air Turbulence with Structure Function

In this section, measurements using a laser Fizeau interferometer are discussed. As Figure 5.1 shows, the laser Fizeau interferometer includes three functional sub-systems: the illumination system with spherical wave (or plane wave) illuminating reference surface at normal incidence, the interference cavity that encodes difference between reference and part under test, and the imaging system that reflects part under test onto detector [113].

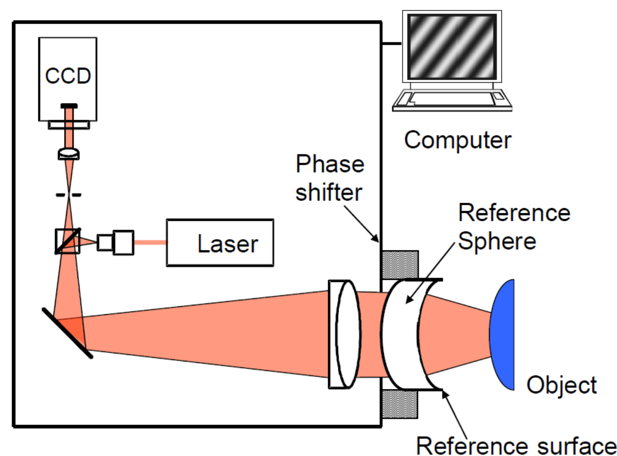


Figure 5.1 Laser Fizeau interferometer configuration [113].

According to the Guide to the Expression of Uncertainty in Measurement (GUM), it is obligatory to give the quantitative indication of the quality of the measurement result of a physical quantity, that is, for evaluating and expressing its uncertainty. Without this, measurement results cannot be compared, either among themselves or with reference values indicated in a specification or standard [114].

In the measurement using the Fizeau interferometer, various noise sources can affect the quality of the result. In this case, we analyzed the uncertainty of two typical sources (electronic noise and air turbulence) by taking a number of independent measurements and computing the averaged SF of the difference between each two adjacent measurements.

5.1.1 Uncertainty Analysis for Electronic Noise by Linear SF

In a temperature controlled lab, a Zygo Verifire Fizeau interferometer was set on a vibration isolated table. As Figure 5.2 shows, the interference cavity length was 5 mm and was insulated with a bubble wrap. This minimizes environmental effects such as the vibration, the temperature, and air turbulence in such a short cavity.

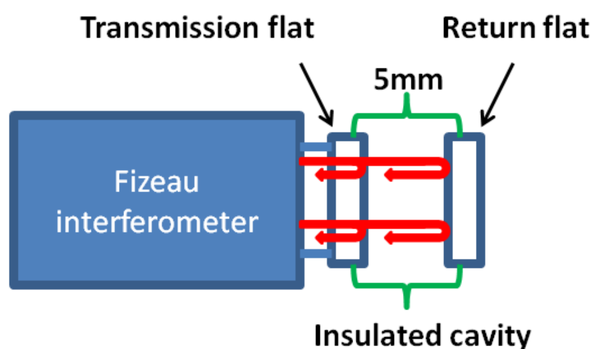
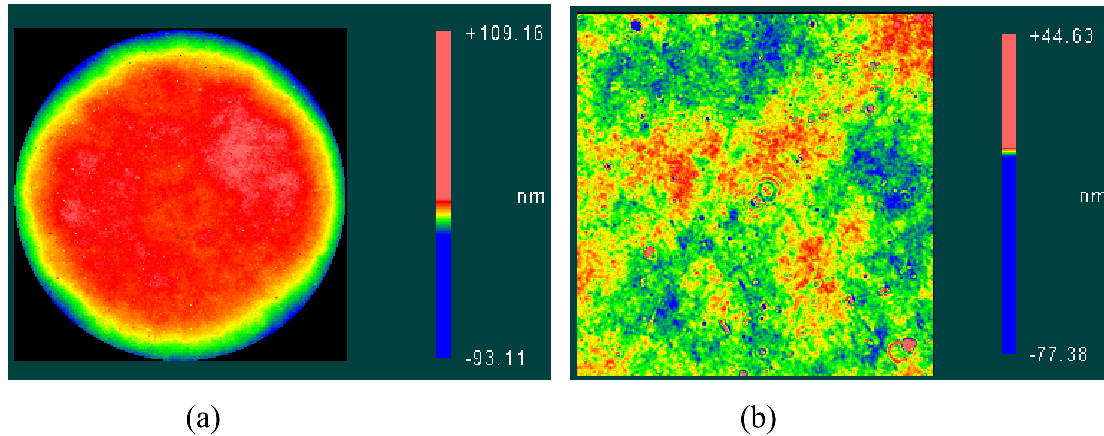


Figure 5.2 Setup of the interferometer with a short insulated cavity.

In this case, the measurements were completed in four groups without changing the setup of the instrument. The data in each group were obtained with the average number of

4, 8, 16, and 32, respectively. Figure 5.3 (a) is the measured surface with average number 32 (Fizeau 1x). Figure 5.3 (b) is the same surface measured with average number 32 when the Fizeau interferometer is zoomed in from 1x to 6x.



(a) (b)

Figure 5.3 Measurement with Fizeau interferometer. (a) Surface data measured with Fizeau 1x (100mm×100mm). (b) Surface data measured with Fizeau 6x (16.7mm×16.7mm)

With each average number (4, 8, 16, 32), ten measurements (M_1, M_2, ... M_{10}) were made. After calculating the SF of the difference between each two adjacent measurements (SF(M_2-M_1), ...), the average was computed. The procedure is:

Group 1: Avg. No. = 4: M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, M_{10}.

SF of difference = [SF(M_2-M_1) + SF(M_3-M_2) + … + SF(M_{10}-M_9)] / 9.

Group 2: Avg. No. = 8: M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, M_{10}.

SF of difference = [SF(M_2-M_1) + SF(M_3-M_2) + … + SF(M_{10}-M_9)] / 9.

Group 3: Avg. No. = 16: M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, M_{10}.

SF of difference = [SF(M_2-M_1) + SF(M_3-M_2) + … + SF(M_{10}-M_9)] / 9.

Group 4: Avg. No. = 32: M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, M_{10}.

SF of difference = [SF(M_2-M_1) + SF(M_3-M_2) + … + SF(M_{10}-M_9)] / 9.

As Figure 5.4 shows, the SF (representing the errors) decreases when the average number increases. And the SF increases when the Fizeau interferometer is zoomed in from 1x to 6x. It may be because the intensity distribution in each pixel increases when zooming the interferometer.

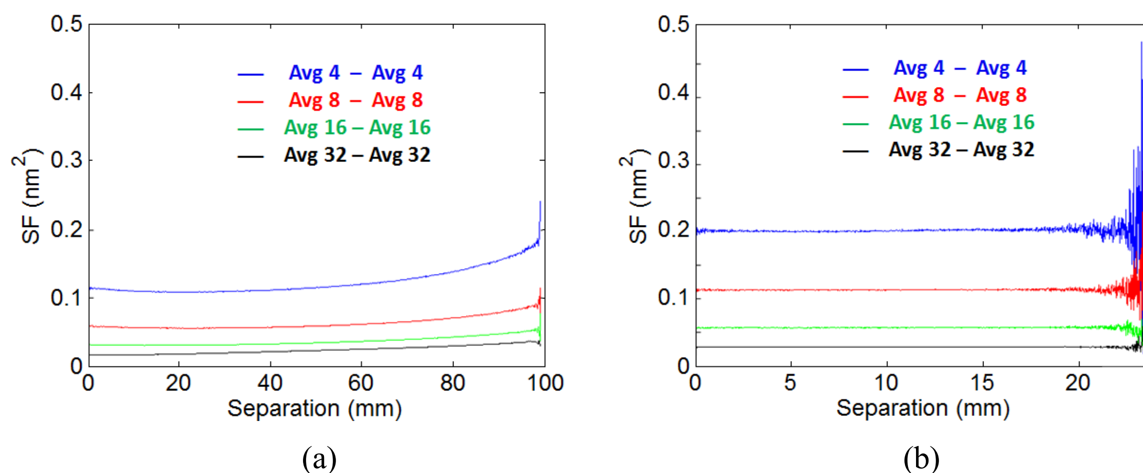
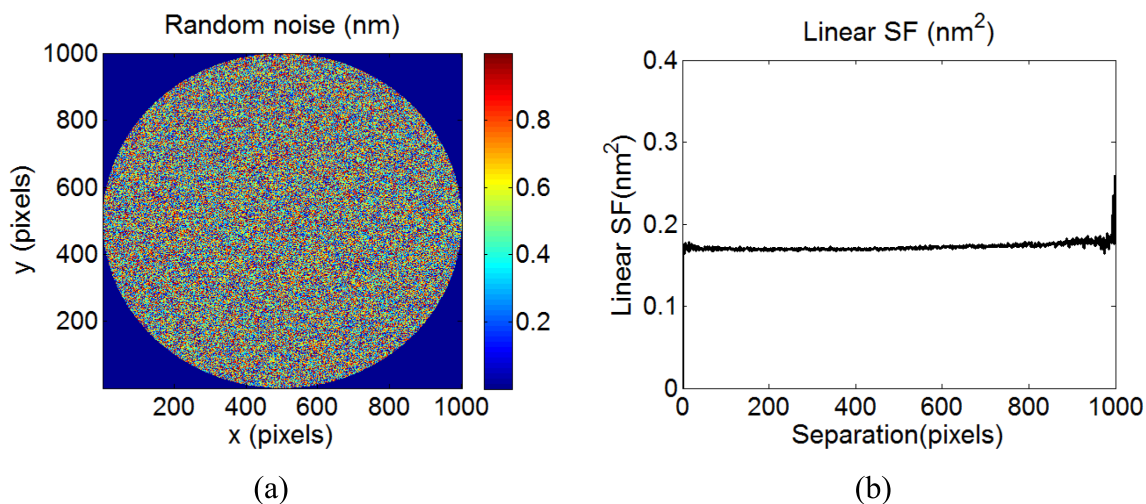


Figure 5.4 SF of the difference. (a) SF of the difference (Fizeau 1x). (b) SF of the difference (Fizeau 6x).

For the 1x data it is noticeable that there is some curvature in the linear SF which decreases with increased averaging. This may be because the intensity distribution decreases, and hence signal to noise ratio increases, toward the edge of the aperture. For the 6x data, it is obvious that most part of the SF is flat (except at separation where there are only a few calculated points). Over the central portion of the aperture, the intensity distribution is almost uniform.



(a) (b)

Figure 5.5 SF analysis for random noise. (a) Random noise. (b) SF of random noise.

Figure 5.5 (a) is a simulated surface with random noise, and Figure 5.5 (b) shows that the SF of the random noise is essentially be flat except that at large separations, due to the relatively small number of calculated points.

Therefore, when comparing Figure 5.4 to Figure 5.5, it is clear that the dominant effect in this measurement is the randomly distributed electronic noise.

5.1.2 Uncertainty Analysis for Air Turbulence by Linear SF

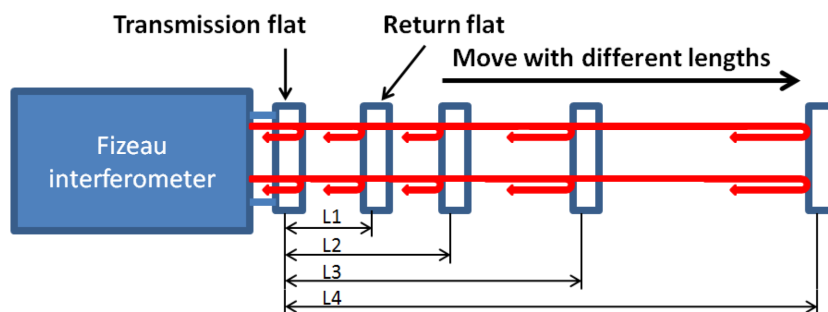


Figure 5.6 Setup of the interferometer with different cavity lengths.

In the same environment as last experiment (temperature controlled lab and vibration isolation table), the Fizeau interferometer was set with different interference cavity lengths (The return flat (RF) was moved to 35 mm, 60 mm, 110 mm, and 210 mm away from the

transmission flat, respectively). In this case, the major source of the environmental effects was the air turbulence.

In each cavity, ten measurements (M_1, M_2, ... M_{10}) were made. After calculating the SF of the difference between each two adjacent measurements (SF(M_2-M_1), ...), the average was computed. The procedure is similar to last experiment.

Cavity 1: Measurement: $M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, M_{10}$.

SF of difference = [SF(M_2-M_1) + SF(M_3-M_2) + … + SF(M_{10}-M_9)] / 9.

Cavity2: Measurement: $M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, M_{10}$.

SF of difference = [SF(M_2-M_1) + SF(M_3-M_2) + … + SF(M_{10}-M_9)] / 9.

Cavity3: Measurement: $M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, M_{10}$.

SF of difference = [SF(M_2-M_1) + SF(M_3-M_2) + … + SF(M_{10}-M_9)] / 9.

Cavity4: Measurement: $M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, M_{10}$.

SF of difference = [SF(M_2-M_1) + SF(M_3-M_2) + … + SF(M_{10}-M_9)] / 9.

Figure 5.7 shows the SFs of the differences for Fizeau 1x (100 mm diameter) and Fizeau 6x (16 mm × 16 mm). Comparison between Figure 5.7 and Figure 5.4 indicates that the environmental effect (air turbulence) is more than an order of magnitude greater than the electronic noise.

As Figure 5.7 shows, the SF (air turbulence) increases when the cavity length increases. Thus, in order to reduce the influence of the air turbulence, it is important to use a short cavity for the Fizeau interferometer, where possible.

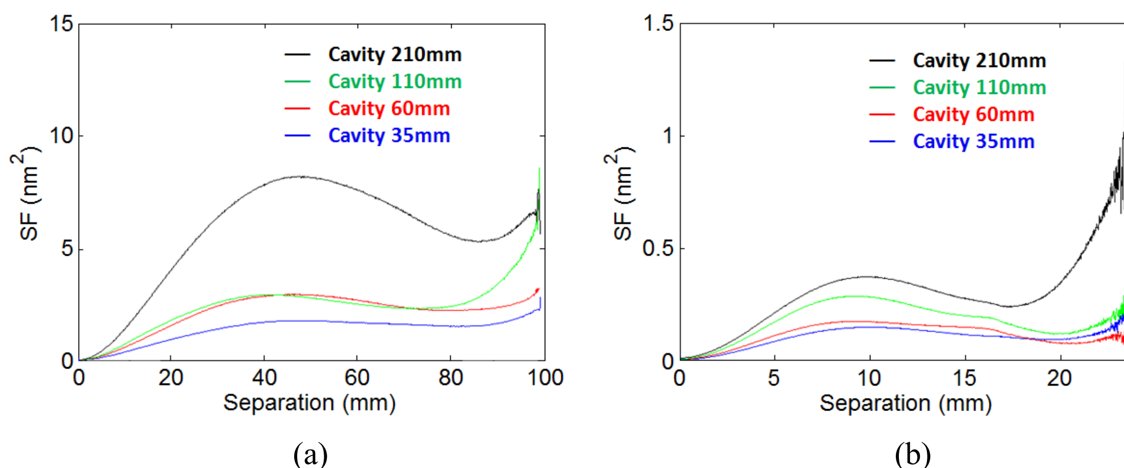


Figure 5.7 SFs of the differences. (a) SF of the difference (Fizeau 1x). (b) SF of the difference (Fizeau 6x).

When comparing Figure 5.7 (a) to (b), it is obvious that the SF decreases when the Fizeau interferometer is zoomed in from 1x to 6x. That is, the influence of air turbulence is related to the size of the dimensions of the average turbulence "cell" sampled by the measurement which decreases when the measuring area decreases.

Moreover, in Figure 5.7, the SF is small on the high frequency domain, because the separation is so small (pixel to pixel) that the influence of air turbulence can be negligible, consequently the dominant noise is the electronic noise.

Figure 5.7 (b) is based on a square aperture (16 mm × 16 mm), then the SFs in the largest separations (near the 24 mm) are calculated from very limited data points and are not meaningful.

Since the air turbulence is also related to thermal gradient (affecting the air refractive index [115,116,117,118]), it is necessary to investigate its distribution in the interference cavity.

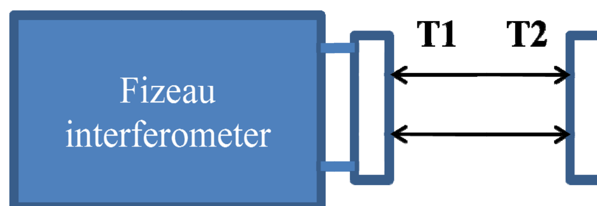


Figure 5.8 Turbulence parallel distribution (introducing piston).

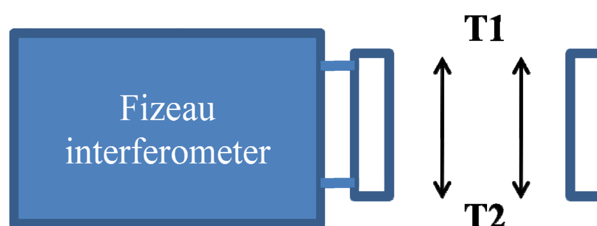


Figure 5.9 Turbulence vertical distribution (introducing tilt).

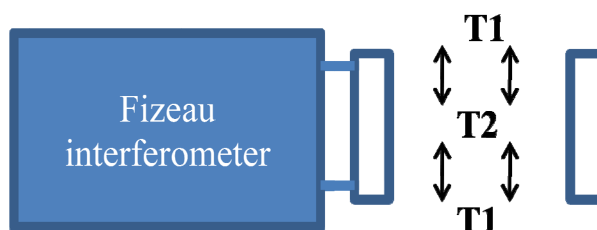


Figure 5.10 Turbulence radial distribution (introducing power).

As Figure 5.8, Figure 5.9 and Figure 5.10 show, there are three distributions: the parallel distribution which introduces the piston, the vertical distribution which introduces the tilt, and the radial distribution which introduces the power.

Due to the impact of the alignment uncertainty, the piston and tilt have been removed from the measurement. Therefore, the power introduced by the air turbulence may play an important role in the SF. Figure 5.11 shows the SF of the simulated power surface, which is similar to the SF in Figure 5.7. Consequently, the large value of the SF in the mid spatial frequency domain may arise from the radial distributed air refractive index in the interference cavity, i.e. approximately spherical turbulence cells.

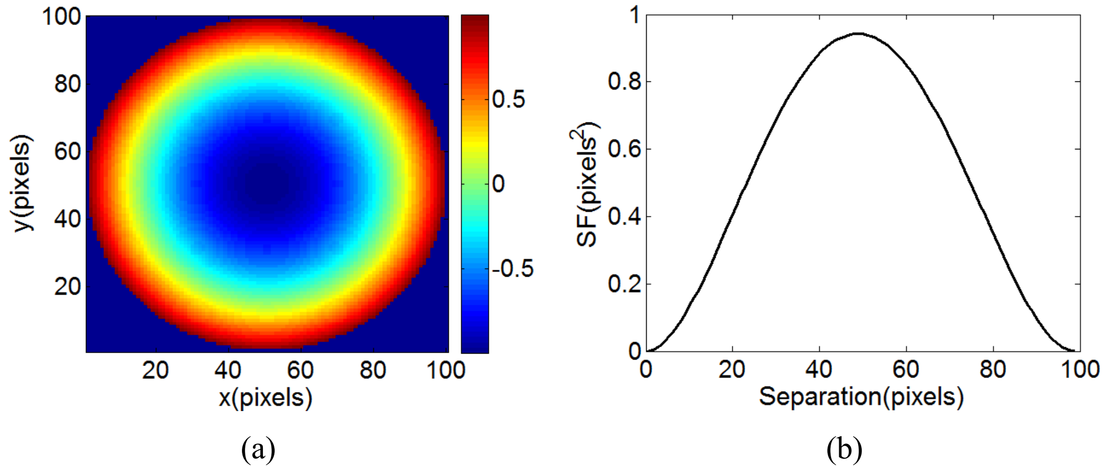


Figure 5.11 Linear SF of the power. (a) Power. (b) Linear SF of the power.

5.1.3 Uncertainty Analysis for Air Turbulence by Area SF

The area SF can also be used to analyze the air turbulence. In this case, a Zygo DynaFiz interferometer is used (Figure 5.12). It is a phase measuring interferometer designed for performing accurate measurement in the environment of air turbulence and extreme vibrations [119].

Figure 5.12 Zygo DynaFiz interferometer [119].

Different from the conventional time-based PSI (for example: Zygo Fizeau interferometer), the DynaFiz instantaneous interferometer is based on the carrier fringe method. It transfers the phase shift to the interference signal which can be measured within a time window of a single shuttered camera frame (for example: 1msec) [120,

121]. Therefore, it is possible to capture the dynamically-changing events ("freeze" the turbulence) with this method.

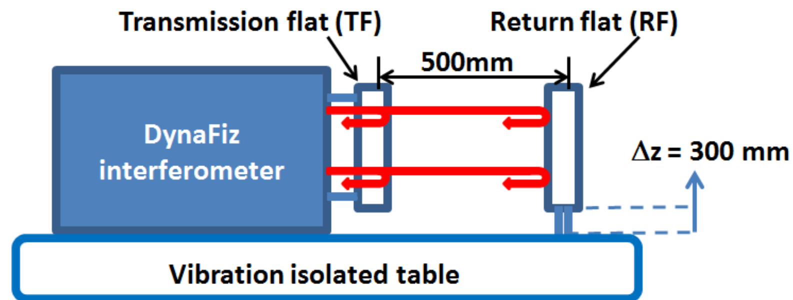


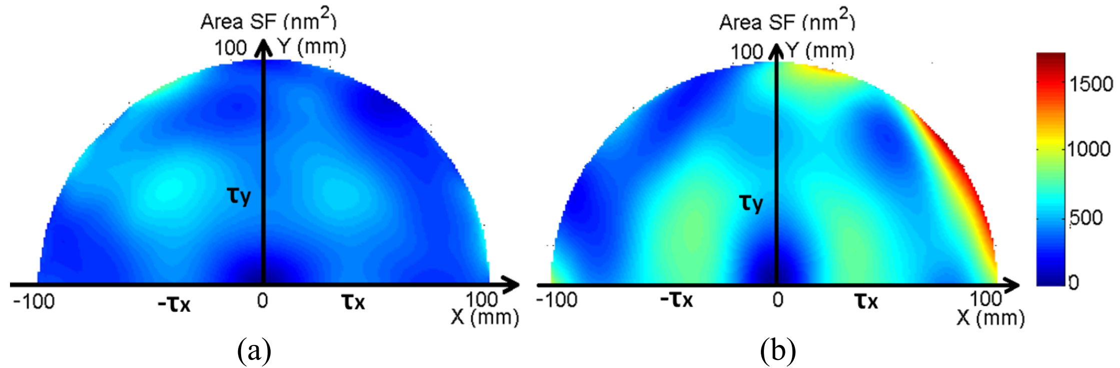
Figure 5.13 Setup of the DynaFiz interferometer.

The experiment was conducted in a temperature controlled lab. As Figure 5.13 shows, the return flat is set 500 mm away from the transmission flat. In this case, the major source of the environmental effects was the air turbulence. The camera shutter of the DynaFiz interferometer was set to 5% of the frame rate, which could reduce the exposure time to freeze the turbulence. Consequently, the frame exposure time was 0.983 msec.

In this experiment, ten measurements were made. After calculating the area SF of the difference between each two adjacent measurements, the average was computed in Figure 5.14 (a).

Then, we lifted up the setup by 300 mm, and made the same measurements. The averaged area SF is shown in Figure 5.14 (b). Because we measured the same transmission and return flats with the same instrument in these two experiments, the only thing that introduced the difference between the two area SFs was the air turbulence. Comparison between Figure 5.14 (a) and (b) denotes that the air turbulence in the high cavity is stronger than that in the low cavity. This may be because of the stagnation at the table top. In both cavities, the air turbulence has more influences in the direction and

separation of ($-\tau_x$, τ_y) and (τ_x, τ_y). In addition to the information of linear SF, the area SF indicates the direction of the air turbulence.



(a) (b)

Figure 5.14 Area SF of turbulence. (a) Area SF of turbulence in low cavity. (b) Area SF of turbulence in high cavity.

5.2 Dynamic Measurement on Diamond Turning Machine

In the precision machining, it is possible to improve workpiece dimensional accuracy by correcting for repeatable errors of the machine tool. Usually the workpiece needs to be removed from the machine and taken to a metrology lab for measurement. After the measurement, the workpiece is set up to the machine again for the compensation machining.

In this case, since the machining and the measurement are separated, it is easy to introduce repositioning errors. One of the solutions to this problem is on-machine measurement [122, 123]. In the case of optics manufacturing, the environment of the machine shop is not usually suitable for traditional interferometry.

Since the Zygo DynaFiz interferometer can implement dynamic acquisition in bad environment, it was chosen to perform the on-machine measurement for diamond turning and the SF was used to analyze the surface in spatial domains.

As Figure 5.15 shows, a workpiece is fixed to the spindle (be able to move in x direction) of a Precitech diamond turning machine (DTM), a mirror is fixed to the precision stage which can move in z direction, and the DynaFiz interferometer is set up on a hydraulic cart beside the DTM. Then, the workpiece surface can be measured by adjusting the x and z axes, the position of the DynaFiz interferometer, the tilt of the mirror, etc.

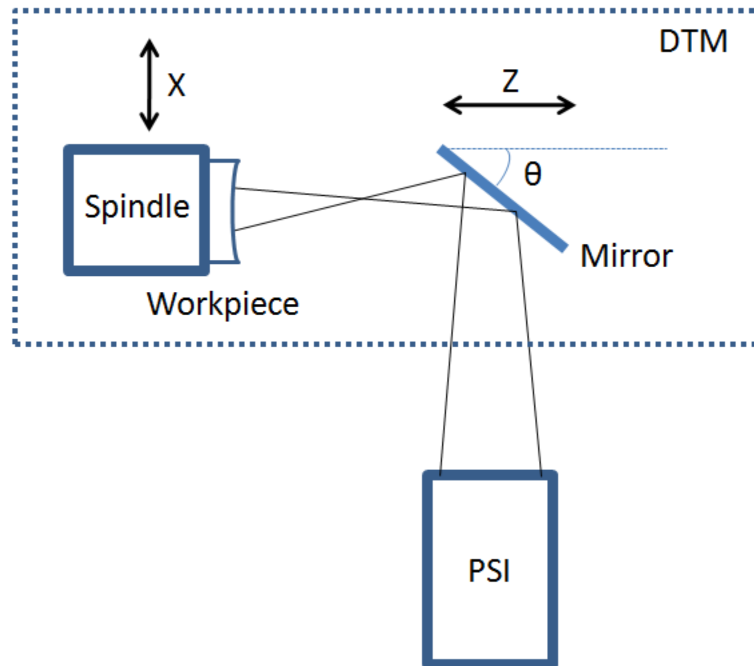


Figure 5.15 Measurement configuration.

In this experiment a PMMA part was single point diamond machined to give a concave spherical surface with an effective diameter of 40 mm and a radius of curvature of 200 mm. Based on the part dimensions and the positioning ranges of the slides in the DTM, a transmission sphere with a diameter of 100 mm and a radius of curvature of 350 mm was chosen for the DynaFiz interferometer.

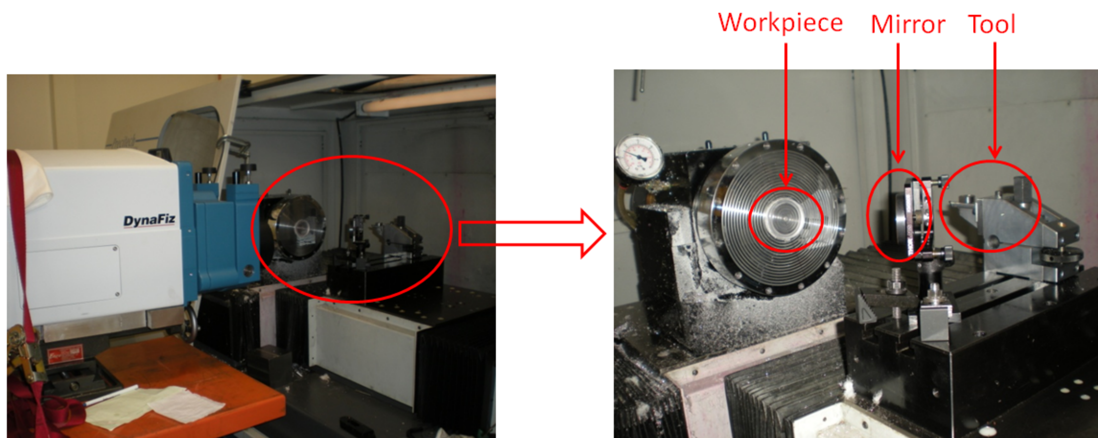


Figure 5.16 Experiment setup after the first machining.

Figure 5.16 shows the setup of the measurement after the first machining. It should be noted that the height data obtained by the DynaFiz interferometer are the departure of the measured spherical surface from a best fit spherical surface.

When the interferometer completes the measurement of this spherical part, it is necessary to remove the tilt from the height data, because it is impacted by alignment uncertainty. Moreover, it is also necessary to remove the power since it is not possible to tell if the part is in the wrong place or the radius of curvature is incorrect.

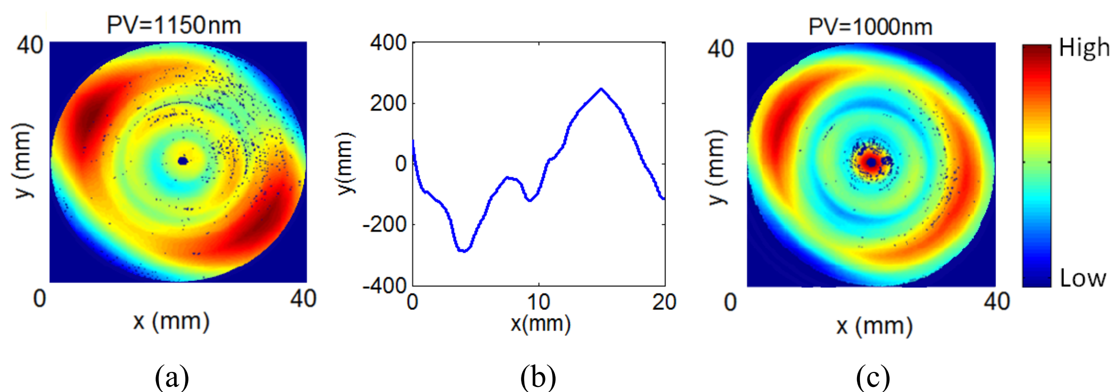
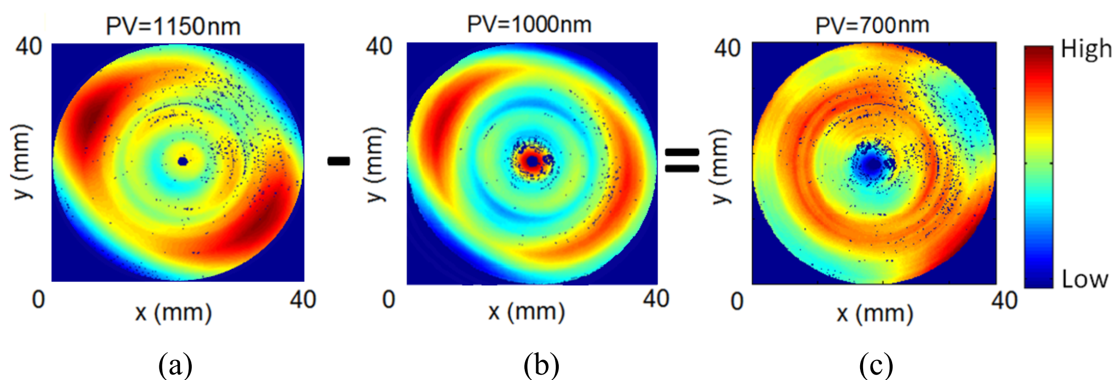


Figure 5.17 Measured height maps. (a) Height map measured after the initial machining. (b) Averaged radial profile. (c) Height map measured after the compensation machining.

Figure 5.17 (a) shows the height map measured after the initial machining. As discussed above, the piston, tilt and power of the raw data have already been removed. With this height map, we took 10 different radial profiles, interpolated to get a common data spacing, and averaged. Based on these averaged radial data (Figure 5.17 (b)), the compensated machine code was generated for the DTM. A height map measured after the compensation machining is shown in Figure 5.17 (c).



(a) (b) (c)

Figure 5.18 Measured height maps. (a) Height map measured after the initial machining. (b) Height map measured after the compensation machining. (c) Difference between the two maps.

Figure 5.18 shows the subtraction from the initial machined surface to the corrected surface. It is clear that the initial part has rotationally varying (astigmatism) and rotationally invariant components (Figure 5.18 (a)). The astigmatism is not likely caused by a machine error, and more likely the result of residual stresses in the sample and mount-induced deformations. After the compensation machining, the astigmatism is still existed, although the height values have been reduced (Figure 5.18 (b)). That is because the diamond turning machine only reduces the rotationally invariant error, in this case including the average rotationally varying error.

Figure 5.18 (c) is the difference between the initial machined surface and the corrected surface. Note that the dominant part is rotationally invariant. However, there are some

rotationally varying components. For example, the high edge at the lower right. It is due to the high value of the astigmatism in this position. During the diamond turning, in the positions with the same radius, the high value can be reduced, while the low value may not be affected.

Because the analysis of the form error (e.g. astigmatism) is not enough for the deterministic manufacturing process, the SF is used to estimate the full spatial frequency component.

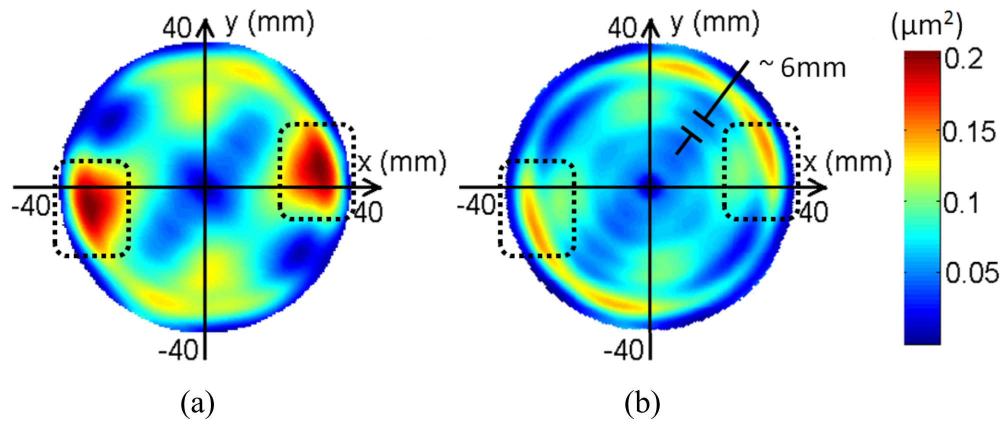


Figure 5.19 Area SF in four quadrants. (a) Initial machining. (b) Compensation machining.

Figure 5.19 shows the area SFs of the height maps in Figure 5.17. After the compensation machining, the SF decreases (especially in the dashed region which is at the large separation of approximately 30 mm) and the distribution of the SF on the entire surface becomes even. This low spatial frequency analysis is consistent with the form error interpretation in Figure 5.18.

Note that the “ripple” in the area SF with a separation repeat length of about 6 mm in the upper-right and lower-left directions, indicating the directional mid spatial frequency

component on the diamond turned surface. However, this characteristic is difficult to be observed directly from the height map in Figure 5.18.

As a result, the on-machine dynamic measurement with the DynaFiz interferometer provides reasonable feedback for the following compensation machining, which improves the accuracy of the workpiece surface. Moreover, the analysis with SF gives a quantitative estimation in the spatial frequency domain.

5.3 Summary

This section investigates the application of the SF. It can be used to estimate various noises for the PSI measurement. As the results show, for the precision measurement, the air turbulence has more influence than the electronic noise of the interferometer, and the long interference cavity introduces more environmental effect than the short cavity. Particularly, an instantaneous interferometer was used to make dynamic measurement of the air turbulence.

In addition, the instantaneous interferometer was applied to implement an on-machine measurement for a diamond turning machine, which provides an effective feedback for the compensation machining. Finally, the analysis with SF gives a quantitative estimation in the spatial domains – the dimensional accuracy of the diamond turned surface has been improved after the compensation machining.

CHAPTER 6 : CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

The structure function (SF) is a useful tool to specify and characterize the spatial content of optical surfaces.

The SF is the squared expectation of the height difference as a function of separation. A linear SF can be computed for profile or area data, with the loss of any detail of anisotropy information. The area SF can be calculated without filtering, zero padding and restraints on aperture size or shape and retains information on anisotropy.

For large numbers of points within the aperture, computational time of the SF increases rapidly, so sampling is required. Comparisons show that the sliding sampling we developed is faster and more accurate than the conventional sampling strategies. For stationary surfaces, the area SF with positive values of separation in x and y is sufficient. However, for non-stationary surfaces, typical of optical surfaces when form is included, a second quadrant is required to fully describe the spatial content of the surface and its anisotropy. The relationship between the linear SF and the two-quadrant area SF has been investigated for circular apertures, showing that the linear SF is the average over concentric semi-circles of the two-quadrant area SF.

The SF has some advantages when compared to other representations. The SF contains similar information as autocorrelation function (ACF) for stationary surface, but the SF provides better visualization of the surface characteristics, since the ACF is strongly

dependent on the mean plane and is not intuitive. The SF is computationally "correct" for any arbitrary aperture shape without extra processing, while the power spectral density (PSD) always needs additional mathematical processing, such as windowing, zero padding, low-order terms removal, and choice of sub-apertures for irregular apertures. There is no standardization of the mathematical treatment of PSDs for optical surfaces; different metrologists may compute different PSDs from the same starting data.

Zernike polynomials are commonly used to characterize and specify circular apertures on optical surfaces. Note, some optical specifications use a SF, but allow for removal of specified amounts of power, coma and astigmatism, etc (defined as combinations of Zernikes). Therefore, it is useful to explore the total SF for a surface described by Zernike polynomials. According to the orthogonality of the Zernike polynomials, the total linear SF is the sum of the individual linear SFs of each of the Zernike polynomials with different azimuthal frequencies. However, this theorem does not apply to the area SF, because each area SF $S(r,\varphi)$ corresponds to only one direction φ, which means the orthogonality cannot be used.

The connection between SF and root mean square (RMS) gradient has been made.

The spatial content of surfaces over large dynamic ranges for arbitrary aperture shapes can be represented by an area SF constructed from data obtained from more than one instrument. Proper combination of area SFs from different instruments requires accounting for the effect of low order form on small separation SF values. Generally speaking, the area SF of the measured sub-aperture (representing the high spatial frequency content) depends on the measurement position in the full aperture, especially for the anisotropic surfaces. To solve this problem, a sampling method of sub-aperture measurement has been

developed. The correct area SF can be estimated by adding the correct tilt to each sampled sub-aperture, typically estimated from the large-area measurement with knowledge of the global coordinates of each sub-aperture measurement. Moreover, the effect of the instrument transfer function (ITF) of the instruments used must also be considered.

The SF can be used to estimate various noise sources for the phase shifting interferometry (PSI). The results confirm previous experience that, for the precision measurement, the air turbulence has more influence than the electronic noise of the interferometer, and the long interference cavity introduces more environmental effect than the short cavity.

An instantaneous interferometer was used to implement on-machine measurement on a diamond turning machine, which provides an effective feedback for the compensation machining. An analysis with SF gives a quantitative estimation – in the spatial domain – of the dimensional accuracy of the diamond turned surface after the compensation machining.

6.2 Future Work

6.2.1 Analysis of Type B Uncertainty for SF Combination

When looking at the combination of the SF from multiple instruments, within the overlapped region in the response (spatial terms) of two instruments, there is often a discrepancy between instruments greater than the noise in the measurements. Chapter 4 has discussed the Type A uncertainty for this discrepancy. A full uncertainty analysis should include Type B uncertainty, such as the effect of an uncalibrated reference surface of the objective, distortion and retrace error. Typically, the full uncertainty evaluation of an area SF will be a matrix with the same size.

In short, doing a complete uncertainty evaluation in the area SF is a nontrivial problem, that is the direction for the future research.

6.2.2 Extend the SF Combination from Flats to Spherical and Freeform Surfaces

The combination of SFs from multiple instruments in this work is only focused on optical flats. It is necessary to extend the analysis to spherical, aspheric, and freeform surfaces.

The first step for any analysis is always a background subtraction to establish the measurement result as the departure of the surface from the intended geometry. This must be done to look at surface errors over any spatial frequency range.

With flat or spherical surfaces, this step is straightforward. The method introduced in Chapter 4 can be easily applied to spherical surfaces. For each sub-aperture measurement of the spherical surface, the x/y translation and the rotation of the measuring stage should be recorded. In other words, the position of the sub-aperture is the combination of the x/y translation and the rotation, which means it is easy to introduce the positioning error. Since error in the estimate of the sub-aperture position leads to error in the tilt correction and therefore error in the SF calculation, it is necessary to use an instrument with high positioning accuracy (e.g. a SWLI with calibrated scales of x and y translation and stage rotation). Besides the improvement of the instrument, other operations and algorithms should be explored to reduce the positioning error.

For freeform surfaces, the measurement is more complicated and the combination of the SF from multiple instruments is more challenging. A freeform surface may not even have an axis of symmetry. Care must be taken in the design, fixturing and fabrication

stages in freeform manufacturing to clearly define a part coordinate system that will allow meaningful subtraction of the design geometry.

Besides figure error and roughness, the mid-spatial error is also easily introduced by the small-tool fabrication for freeform surfaces. Consequently, it is nontrivial to investigate the SF combination for freeform surfaces, which can characterize the full spatial frequency content.

6.2.3 Application of Two-quadrant Area SF in Astronomical Optics Specification

As noted in Chapter 1, in order to make sure that the astronomical telescope performance degradation arising from the optical errors is less than the best atmosphere from the statistical analysis, the linear SF has been used to characterize the phase errors introduced by the atmospheric turbulence on various spatial scales. Since the area SF has some advantages compared to the linear SF, for example that it retains of anisotropic information of the surface, it has potential application to specify the astronomical optics in spatial domain. This section demonstrates two simulations to discuss the application of the two-quadrant area SF for Thirty Meter Telescope (TMT) segment.

6.2.3.1 Linear and Area SF Specification of the TMT Segment

Based on the atmospheric analysis, the linear SF specification of the TMT segment surface figure accuracy is defined as [64]:

$$S(\tau) = A[10.60(\tau/d)^{5/3} - 13.75(\tau/d)^{2} + 3.42(\tau/d)^{3}] + 2B^{2}. \quad (6.1)$$

Where:

$S(\tau)$ is the SF in nanometers squared

$$A = \left(\frac{1}{2}\right)^{2}\left(\frac{500nm}{2\pi}\right)^{2}\left(\frac{d}{r_0}\right)^{5/3}$$

A = leading coefficient = 2907 nm^2

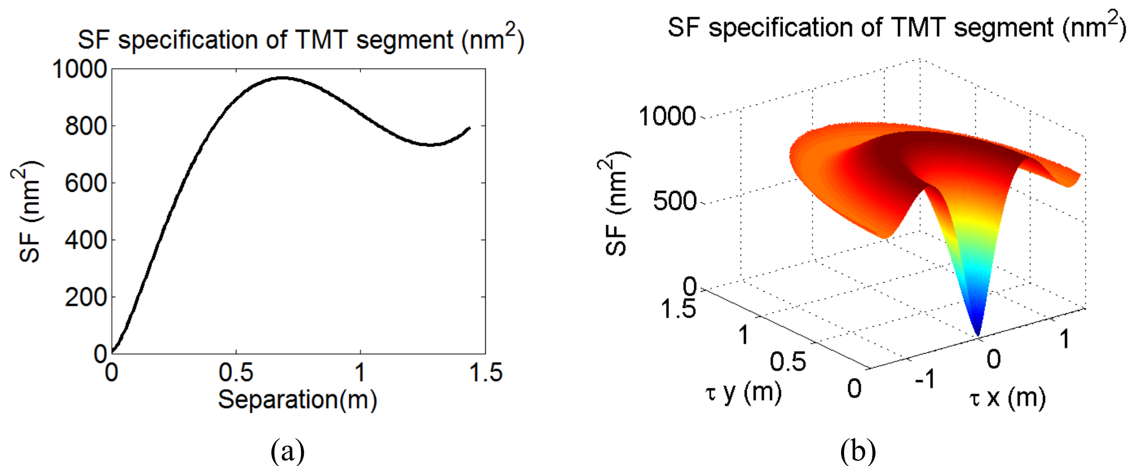
B = high frequency surface roughness = 2 nm

τ = separation between point pairs

d = diameter of segment = 1.44 m

r_0 = Quasi-Fried's parameter = 1.0 m

This SF specification is in terms of surface errors (not wavefront) with piston and tilt removed from the phase map [64]. The curve is shown in Figure 6.1 (a). Thus, the SF of the measured surface errors (including both the measured aberrations and the measurement uncertainty) shall be less than the specification at all separation distances (spatial frequencies).



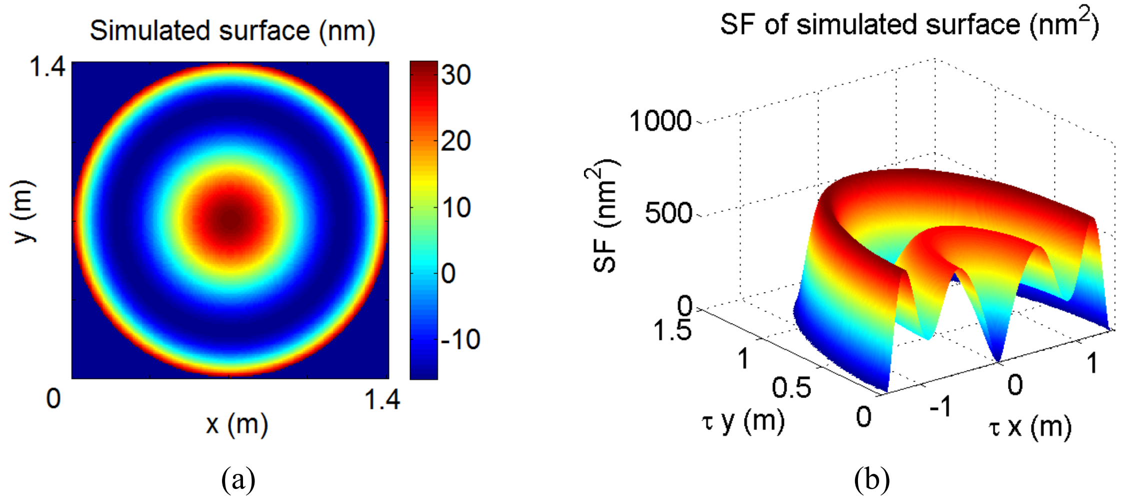
(a) (b)

Figure 6.1 SF specification of TMT segment surface figure accuracy. (a) Linear SF specification of TMT segment. (b) Rotate the linear SF as an area SF.

It can be assumed that the two-quadrant area SF specification is obtained by rotating the linear SF with 180° (Figure 6.1 (b)). In fact, the TMT segment is a hexagonal mirror. For simplicity, it is treated as a circular shape in this case. Because the calculation of the linear SF is based on the diameter of the hexagonal segment, it is reasonable to use the circular aperture with this diameter.

To investigate the application of the area SF for the TMT segment, two simulated surface errors are discussed in the following sections.

6.2.3.2 Simulation NO. 1



(a) (b)

Figure 6.2 SF of the simulated surface error. (a) Simulated surface error. (b) Area SF of simulated surface error.

Figure 6.2 (a) is a simulated rotationally invariant surface error of the TMT segment, and Figure 6.2 (b) is its two-quadrant area SF. In order to know if the simulated segment meets the requirement of the surface figure accuracy, it is necessary to compare the calculated area SF to the specification.

As Figure 6.3 (a) shows, the two-quadrant area SFs of both the simulated surface error and the specification are plotted on the same scale. It is clear that all the simulated SFs are below the specification, which means the simulated part meet the requirement. Figure 6.3 (b) is the linear SF analysis, which shows the same result as area SF.

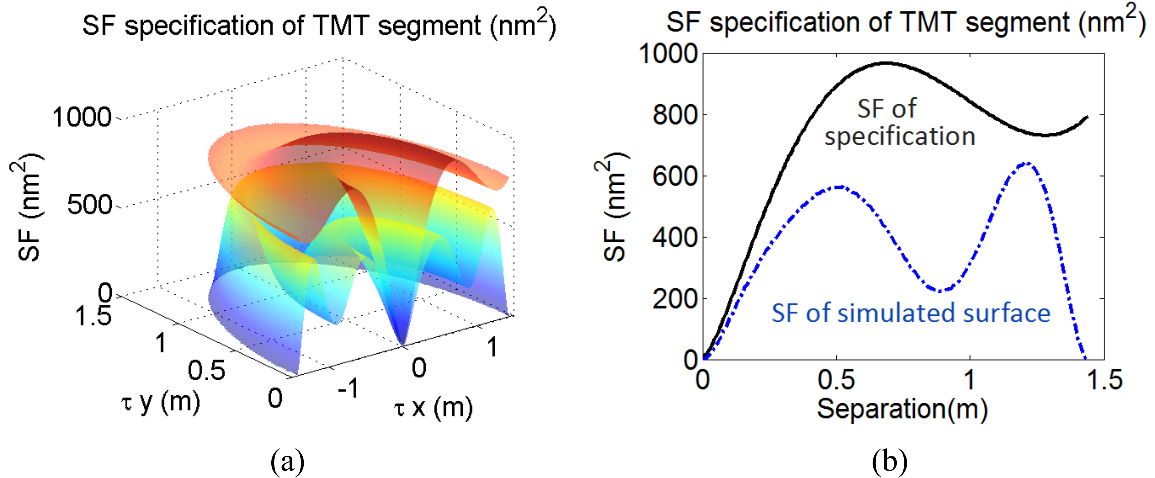


Figure 6.3 SF analysis with simulated surface error. (a) Area SF analysis. (b) Linear SF analysis.

6.2.3.3 Simulation NO. 2

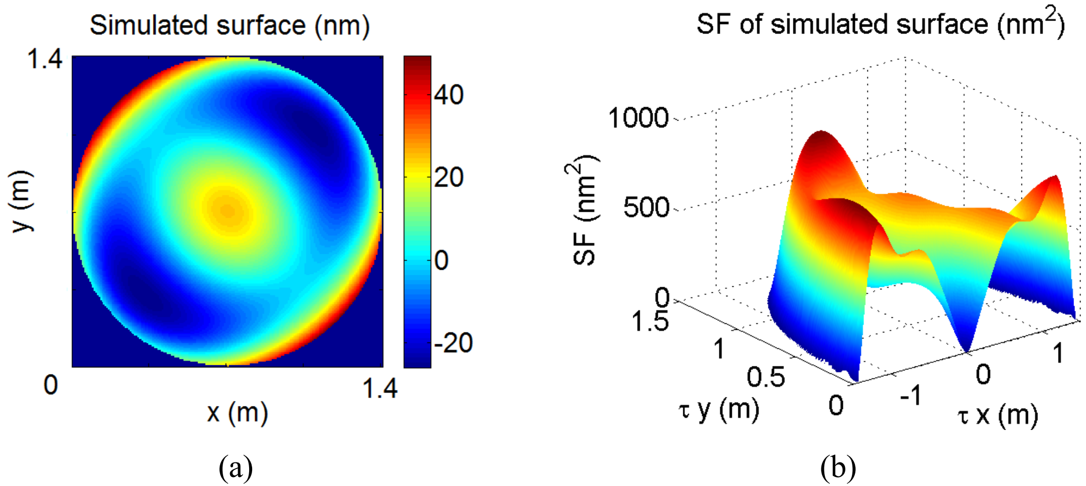


Figure 6.4 SF of the simulated surface. (a) Simulated surface error. (b) Area SF of simulated surface error.

Figure 6.4 (a) is a simulated rotationally variant surface error of the TMT segment, and Figure 6.4 (b) is its two-quadrant area SF. Note that the area SF is also rotationally varying.

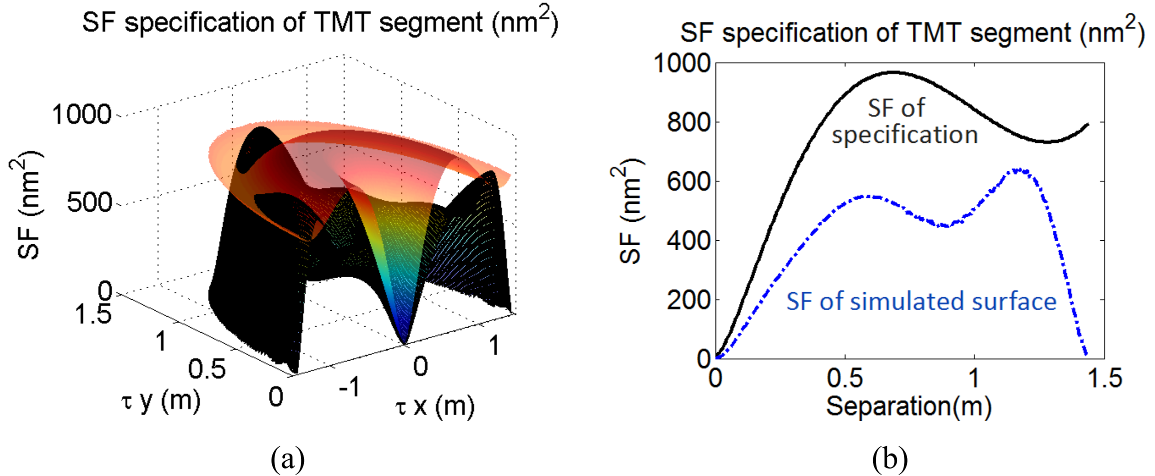


Figure 6.5 SF analysis with simulated surface error. (a) Area SF analysis. (b) Linear SF analysis.

When plotting the two-quadrant area SFs of both the simulated surface error and the specification on the same scale (Figure 6.5 (a)), it is obvious that some peaks of the simulated area SF exceed the specification. That is, the simulated part does not meet the requirement. However, the analysis with the linear SF in Figure 6.5 (b) shows the simulated linear SF is below the specification, which means the simulated part meet the requirement. In brief, the simulated part passes the linear SF but fails the area SF. The reason is that the linear SF is the average over concentric semi-circles of the two-quadrant area SF. The impact of anisotropy has been reduced in the linear SF.

Thus, when the surface error is rotationally varying, it is necessary to use the two-quadrant area SF as the specification. The analysis with linear SF could result in poor performance of the optics if a bad part is accepted because of the loss of the anisotropic specification.

Therefore, in the future work, the two-quadrant area SF can be applied to specify the astronomical optics in spatial domain.

6.2.4 SF in General Optics Specification and Conformance Testing

Since the SF has some advantages compared to PSD, it has potential application in optical specification and conformance testing. Similar to the PSD defined in ISO 10110 - 8, the SF can be used as the specification of general optical surfaces.

Following the default decision rule of ISO 14253-2, the expanded uncertainty in the SF must be subtracted from the specification for a vendor to prove conformance. Reducing the uncertainty increases the manufacturing headroom.

Therefore, the future work should include the investigation of SF in general optics specification and conformance testing.

REFERENCES

[1] T. R. Thomas, *Rough surfaces* (Imperial College Press, London, 1999).

[2] D. J. Whitehouse, *Handbook of surface and nanometrology* (CRC Press, New York, 2010).

[3] J. Peters, J. Bryan, W. T. Estler, C. Evans, et al., “Contribution of CIRP to the development of metrology and surface quality evaluation during the last fifty years,” CIRP Annals **50** (2), 471-488 (2001).

[4] C. A. Brown, W. A. Johnsen, R. M. Butland, J. Bryan, “Scale-sensitive fractal analysis of turned surfaces,” CIRP Annals **45** (1), 515-518 (1996).

[5] X. Jiang, D. J. Whitehouse, “Technological shifts in surface metrology,” CIRP Annals **61** (2), 815-836 (2012).

[6] Y. I. Rabinovich, J. J. Adler, A. Ata, J. K. Singh, and B. M. Moudgil, “Adhesion between nanoscale rough surfaces: II. measurement and comparison with theory,” Journal of Colloid and Interface Science **232** (1), 17-24, 2000.

[7] E. M. V. Hoek and G. K. Agarwal, “Extended DLVO interactions between spherical particles and rough surfaces,” Journal of Colloid and Interface Science **298** (1), 50-58, 2006.

[8] D. M. Aikens, J. E. DeGroote, R. N. Youngworth, “Specification and control of mid-spatial frequency wavefront errors in optical systems,” in Frontiers in Optics 2008 / Laser Science XXIV / Plasmonics and Metamaterials / Optical Fabrication and Testing, OSA Technical Digest (CD) (Optical Society of America, OTuA1, 2008).

[9] C. Evans, “Robust estimation of PV for optical surface specification and testing,” in Frontiers in Optics 2008 / Laser Science XXIV / Plasmonics and Metamaterials / Optical Fabrication and Testing, OSA Technical Digest (CD) (Optical Society of America, OWA4, 2008).

[10] C. J. Evans, “PVr – a robust amplitude parameter for optical surface specification,” Optical Engineering **48**(4), 043605 (2009).

[11] J. K. Lawson, C. R. Wolfe, K. R. Manes, J. B. Trenholme, D. M. Aikens, and R. E. English, “Specification of optical components using the power spectral density function,” Proc. SPIE **2536**, 38-50 (1995).

[12] 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,” Proc. SPIE **2576**, 281-292 (1995).

[13] L. He, C. J. Evans, and A. Davies, “Two-quadrant area structure function analysis for optical surface characterization,” Opt. Express **20** (21), 23275-23280 (2012).

[14] R. N. Youngworth and B. D. Stone, “Simple estimates for the effects of mid-spatial-frequency surface errors on image quality,” Appl. Opt. **39**, 2198-2209 (2000).

[15] J. E. Harvey and K. L. Lewotsky, “Scattering from multilayer coatings: a linear systems model,” Proc. SPIE **1530**, 35-41 (1991).

[16] W. B.Wetherell, “Effects of mirror surface ripple on image quality,” Proc. SPIE **0310**, 77-93 (1981).

[17] J. M. Tamkin and T. D. Milster, “Effects of structured mid-spatial frequency surface errors on image performance,” Appl. Opt. **49**, 6522-6536 (2010).

[18] J. E. DeGroote, A. E. Marino, K. E. Spencer, and S. D. Jacobs, “Power Spectral Density plots inside MRF spots made with a polishing abrasive-free MR fluid,” Proc. SPIE **TD03**, 134-138 (2005).

[19] R. J. Noll, “Effect of mid and high spatial frequencies on optical performance,” Opt. Eng. **18**, 137-142 (1979).

[20] M. Schulz, I. Weingaertner, C. Elster, and J. Gerhardt, “Low-and mid-spatial-frequency component measurement for aspheres,” Proc. SPIE **5188**, 287-295 (2003).

[21] http://www.savvyoptics.com/files/MSF_ripple_presentationSep_26-3.pdf

[22] R. N. Youngworth, B. B. Gallagher, and B. L. Stamper, “An overview of power spectral density (PSD) calculations,” Proc. SPIE **5869**, 58690U (2005).

[23] J. E. Harvey and A. Kotha, “Scattering effects from residual optical fabrication errors,” Proc. SPIE **2576,** 155-174 (1995).

[24] ISO 10110-8 *Optics and optical instruments: preparation of drawings for optical elementsand systems Part 8: Surface texture* (2002).

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

[26] D. J. Whitehouse, *The properties of random surfaces of significance in their contact* (PhD thesis at University of Leicester, 1971).

[27] L. He, C. J. Evans, and A. Davies, “Characterizing optical surfaces using the area structure function,” Imaging and Applied Optics Technical Papers, OSA technical digest,

OTu1D.6 (Optical Society of America, 2012).

[28] L. He, C. J. Evans, and A. Davies, "Spatial content analysis for precision surfaces with the area structure function," Surface Topography: Metrology and Properties (submitted).

[29] A. Davies, L. He, and C. J. Evans, "Mid-spatial frequency specification and characterization for freeform surfaces," OSA freeform optics meeting, Tucson, AZ (2013) (submitted).

[30] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Transactionson Speech and Audio Processing **9**(5), 504-512 (2001).

[31] E. A. Avallone and T. Baumeister III, *Marks' standard handbook for mechanical engineers*, 10th Ed., (McGraw-Hill, New York, 1996).

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

[33] H. H. Toebben, G. A. Ringel, F. Kratz, and D.-R. Schmitt, "Use of Power Spectral Density (PSD) to specify optical surfaces," Proc. SPIE **2775**, 240-250 (1996).

[34] A. Duparré, J. Ferre-Borrull, S. Gliech, G. Notni, J. Steinert, J. Bennett, "Surface characterization techniques for determining the root-mean-square roughness and power spectral densities of optical components," Appl. Opt. **41**, 154-171 (2002).

[35] C. L. Vernold and J. E. Harvey, "Effective surface PSD for bare hot isostatic pressed (HIP) beryllium mirrors", SPIE **1530**, 144-149 (1991).

[36] J. K. Lawson, C. R. Wolfe, K. R. Manes, J. B. Trenholme, D. M. Aikens, and R. E. English, Jr., "Specification of optical components using the power spectral density function," Proc. SPIE **2536**, 38-50 (1995).

[37] J. K. Lawson, D. M. Aikens, R.E. English, Jr., and C.R. Wolfe, "Power spectral density specifications for high-power lasersystems," Proc. SPIE **2775**, 345-356 (1996).

[38] E. L. Church and P. Z. Takacs, "Basic program for power spectrum estimation," Brookhaven National Laboratory Informal Report, BNL 49035, (1993) (revised 5/94).

[39] J. H. Campbell, R. A. Hawley-Fedder, C. J. Stolz, "NIF optical materials and fabrication technologies: an overview," Proc. SPIE **5341**, 84-101 (2004).

[40] L. Blunt and X. Jiang (eds.), *Advanced techniques for assessment of surface topography development of a basis for 3D surface texture standards "SURFSTAND"* (Kogan Page Science, London, 2003).

[41] E. Sidick, "Power spectral density specification and analysis of large optical surfaces," Proc. SPIE **7390**, 73900L (2009).

[42] G. Heinzel, A. Rudiger, and R. Schilling, "Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new flat-top windows," Report, Max Plank Institute, Teilinstitut Hannover, 11-12 (2002).

[43] P. Z. Takacs, "Minimizing artifacts in analysis of surface statistics," Imaging and Applied Optics Technical Papers, OSA Technical Digest, OTu1D.5 (Optical Society of America, 2012).

[44] J. R. Womersley, and M. R. Hopkins, "Suggestions concerning the use of the correlogram for the interpretation of measurements of surface finish," Journees des Etats de surface, 135-139 (1945).

[45] J. Peklenik, "Investigation of the surface typology," CIRP Annals **15**, 381 (1967).

[46] D. J. Whitehouse, *Handbook of surface metrology* (Institute of Physics, Bristol, 1994).

[47] D. J. Whitehouse and J. F. Archard, "Properties of random processes of significance in their contact," ASME Conference on Surface Mechanics, 36-57 (Los Angels 1969).

[48] B. Muralikrishnan and J. Raja, *Computational surface and roundness metrology* (Springer-Verlag, London, 2009).

[49] A. M. Hvisc and J. H. Burge, "Structure function analysis of mirror fabrication and support errors," Proc. SPIE **6671**, 66710A (2007).

[50] V. I. Tatarski, *Wave propagation in a turbulent medium* (McGraw-Hill, New York, 1961).

[51] D. L. Fried, "Statistics of a geometric representation of wavefront distortion," J. Opt. Soc. Am. **55** (11), 1427-1435 (1965).

[52] A. N. Kolmogorov, "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers," C. R. (Doklady) Acad. Sci. URSS **30**, 301-305 (1941).

[53] R. E. Parks, "Specifications: figure and finish are not enough," Proc. SPIE **7071**, 70710B (2008).

[54] D. S. Brown, "Optical specification of ground based telescopes," Proc. SPIE **399**, 12-14 (1983).

[55] R. E. Parks, “Optical specifications and tolerances for large optics,” Proc. SPIE **406**, 98-105 (1983).

[56] J. R. P. Angel, “Designing 8-m Mirrors for the Best Sites,” Identification, Optimization, and Protection of Optical Telescope Sites, ed. R. L. Millis, O. G. Franz, H. D. Ables, and C. C. Dahn, 167-176 (1987).

[57] J. M. Hill, “Optical design, error budget and specifications for the columbus project telescope,” Proc. SPIE **1236**, 86-107 (1990).

[58] J. H. Burge, L. B. Kot, H. M. Martin, R. Zehnder, and C. Zhao, “Design and analysis for interferometric measurements of the GMT primary mirror segments,” Proc. SPIE **6273**, 62730M (2006).

[59] M. Johns, “The Giant Magellan Telescope (GMT),” Proc. SPIE **6267**, 626729 (2006).

[60] H. M. Martin, J. R. P. Angel, J. H. Burge, B. Cuerden, W. B. Davison, M. Johns, J. S. Kingsley, L. B. Kot, R. D. Lutz, S. M. Miller, S. A. Shectman, P. A. Strittmatter and C. Zhao, “Design and manufacture of 8.4 m primary mirror segments and supports for the GMT,” Proc. SPIE **6273**, 62730E (2006).

[61] J. H. Burge, T. Zobrist, L. B. Kot, H. M. Martin and C. Zhao, “Alternate surface measurements for GMT primary mirror segments,” Proc. SPIE **6273**, 62732T (2006).

[62] H. M. Martin, R. G. Allen, J. H. Burge, L. R. Dettmann, D. A. Ketelsen, S. M. Miller and J. M. Sasian, “Fabrication of mirrors for the Magellan Telescopes and the Large Binocular Telescope,” Proc. SPIE **4837**, 609-618 (2003).

[63] J. Nelson and G. H. Sanders, “The status of the Thirty Meter Telescope project,” Proc. SPIE **7012**, 70121A (2008).

[64] Anon. *Specification for finished 1.44-meter primary mirror segments*, TMT. OPT. SPE. 07. 002. CCR03 (2008).

[65] A. Thomas, T. R. Thomas, “Digital analysis of very small scale surface roughness,” J. Wave Mater. Interact. **3**, 341-350 (1988).

[66] J. C. Russ, *Fractal surfaces* (Plenum, New York, 1994).

[67] C. A. Brown and B. Meacham, “Tiling strategies in the patchwork method and the determination of scale area relationships,” Fractals **2**(3) 436-443 (1994).

[68] T. R. Thomas and B. G. Rose´n, “Determination of the optimum sampling interval for rough contact mechanics,” Tribo. Int. **33**, 601-610 (2000).

[69] D. J. Whitehouse, “Some theoretical aspects of structure functions, fractal parameters and related subjects,” Proc. Instn. Mech. Engrs. **215**, 207-210 (2001).

[70] T. R. Thomas, B. G. Rose´n, and N. Amini, “Fractal characterisation of the anisotropy of rough surfaces,” Wear **232**, 41-50 (1999).

[71] T. R. Thomas and B. G. Rose´n, “Surfaces generated by abrasive finishing processes as self-affine fractals,” Int. J. Surf. Sci. Eng. **3**(4), 275-285 (2009).

[72] A. Majumdar and C. L. Tien, “Fractal characterisation and simulation of rough surfaces,” Wear **136**, 313-327 (1990).

[73] B. Bhushan and A. Majumdar, “Elastic-plastic contact model for bifrac-tal surfaces,” Wear **153** 53-64 (1992).

[74] J. A. Greenwood and J. B. P. Williamson, “Contact of nominally flat surfaces,” Proc. Roy. Soc. Lond. A **295**, 300-319 (1966).

[75] B-G.Rosén, R. Ohlsson, and T. R. Thomas, “Wear of cylinder bore microtopography,” Wear **198**, 271-279 (1996).

[76] B.-G. Rosén, R. Ohlsson, T. R. Thomas, “Nano metrology of cylinder bore wear,” International Journal of Machine Tools and Manufacture **38**, 519-527 (1998).

[77] C. Y. Poon, R. S. Sayles and T. A. Jones, “Surface measurement and fractal characterization of naturally fractured rocks,” J. Phys. D: Appl. Phys. **25** 1269-1275 (1992).

[78] R Zhelem, “Specification of optical surface accuracy using the structure function,” Proc. of SPIE **8083**, 808310 (2011).

[79] J. Wyant and K. Creath, “Basic wavefront aberration theory for optical metrology,” in Vol. 11 of Applied Optics and Optical Engineering Series, R. R. Shannon and J. C. Wyant, eds. 1-53 (Academic, New York, 1992).

[80] C-J. Kim and R. R. Shannon, “Catalog of Zernike polynomials,” in Vol. 10 of Applied Optics and Optical Engineering Series, R. R. Shannon and J. C. Wyant, eds. 193-221 (Academic, New York, 1987).

[81] ISO/TR 14999-2, *Optics and Photonics—Interferometric measurement of optical elements and optical systems —Part 2: Measurement and evaluation techniques* (2005).

[82] C. J. Evans, R. E. Parks, P. J. Sullivan, and J. S. Taylor, “Visualization of surface figure by the use of Zernike polynomials,” Appl. Opt. **34,** 7815-7819 (1995).

[83] http://www.loft.optics.arizona.edu

[84] W. Swantner and W.W. Chow, “Gram–Schmidt orthonormalization of Zernike polynomials for general aperture shapes,” Appl. Opt. **33**, 1832–1837 (1994).

[85] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*, 7th ed. (Cambridge University Press, New York, 1999).

[86] D. Malacara, *Optical shop testing*, 3rd ed. (John Wiley & Sons, New Jersey, 2007).

[87] J. Loomis, “Fringe users manual,” in D. Anderson, Fringe Manual Version 3 (Optical Sciences Center, University of Arizona, Tucson, Arizona, 1982).

[88] R. J. Noll, “Zernike polynomials and atmospheric turbulence,” J. Opt. Soc. Am., **66**(3), 207-211 (1976).

[89] V. N. Mahajan, “Zernike circle polynomials and optical aberrations of systems with circular pupils,” Eng. Lab. Notes in Opt. & Phot. News 5 (1994).

[90] J. Y. Wang and D. E. Silva, “Wavefront interpretation with Zernike polynomials,” Appl. Opt. **19**, 1510–1518 (1980).

[91] ASME B46.1, *Surface Texture (Surface Roughness, Waviness and Lay)* (2002).

[92] E. L. Church and P. Z. Takacs, “Surface scattering,” in *Handbook of optics*. vol. I, M. Bass, ed. (McGraw-Hill, New York, 2009).

[93] E. L. Church and P. Z. Takacs, “Specification of surface figure and finish in terms of system performance,” Appl. Opt. **32** (19), 3344-3353 (1993).

[94] http://en.wikipedia.org/wiki/Wiener%E2%80%93Khinchin_theorem

[95] L. He, A. Davies and C. J. Evans, “Comparison of the area structure function to alternate approaches for optical surface characterization,” Proc. SPIE **8493,** 84930C (2012).

[96] R. B. Blackman and J. W. Tuckey, *The measurement of power spectra* (Dover, New York, 1958).

[97]J. M. Bennett and L. Mattsson, *Introduction to surface roughness and scattering*, 2nd ed. (Optical Society of America, Washington, D.C., 1999).

[98] L. He, C. J. Evans, and A. Davies, “Optical surface characterization with the area structure function,” CIRP Annals **62**, 539-542 (2013).

[99] P. de Groot and X. Colonna de Lega, “Interpreting interferometric height measurements using the instrument transfer function,” *Fringe 2005: the 5th International*

Workshop on Automatic Processing of Fringe Patterns, W. Osten, Ed., 30-37 (Springer Verlag, Berlin Heidelberg, 2005).

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

[101] D. M. Sykora and P. de Groot, "Instantaneous measurement Fizeau interferometer with high spatial resolution," Proc. SPIE **8216**, 812610 (2011).

[102] E. Novak, C. Ai, and J. C. Wyant, "Transfer function characterization of laser Fizeau interferometer for high spatial frequency phase measurements," Proc. SPIE **3134**, 114-121 (1997).

[103] C. R. Wolfe, J. D. Downie, and J. K. Lawson, "Measuring the spatial frequency transfer function of phase-measuring interferometers for laser optics," Proc. SPIE **2870**, 553-557 (1996).

[104] V. V. Yashchuk, W. R. McKinney, and P. Z. Takacs, "Binary pseudorandom grating standard for calibration of surface profilometers," Opt. Eng. **47**, 073602 (2008).

[105] S. K. Barber, P. Soldate, E. H. Anderson, R. Cambie, S. Marchesini, W. R. McKinney, P. Z. Takacs, et al., "Binary pseudo-random gratings and arrays for calibration of the modulation transfer function of surface profilometers," Proc. SPIE **7448**, 744803 (2009).

[106] R. Barakat, "Determination of the optical transfer function directly from the edge spread function," J. Opt. Soc. of Am. **55** (10), 1217-1219 (1965).

[107] P. Takacs, M. Li, K. Furenlid, E. Church, "A step-height standard for surface profiler calibration," Proc. SPIE **1995**, 235-244 (1993).

[108] L. He, C. J. Evans, and A. Davies, "Single and multiple-instrument surface characterization with the area structure function," Proc. ASPE **54**, 488-491 (2012).

[109] V. G. Badami, J. Liesener, C. J. Evans, P. de Groot, "Evaluation of the measurement performance of a coherence scanning microscope using roughness specimens," Proc. ASPE **52**, 23-26 (2011).

[110] ISO 14253-2, *Geometrical product specifications (GPS) -- Inspection by measurement of workpieces and measuring equipment -- Part 2: Guidance for the estimation of uncertainty in GPS measurement, in calibration of measuring equipment and in product verification* (2011).

[111] L. L. Deck, C. Evans, "High performance Fizeau and scanning white-light interferometers for mid-spatial frequency optical testing of free-form optics," Proc. SPIE **5921**, 59210A (2005).

[112] D. Sykora, Personal communication with C. Evans, (2013).

[113] C. Evans, Lecture 20, Int. to Opt. Fab & Test, UNC Charlotte courseware, (Spring 2011).

[114] ISO/IEC Guide 98, *Guide to the expression of uncertainty in measurement (GUM)* (1993).

[115] E. Miao and Y. Gu, "Temperature influence on Fizeau interferometer repeatability," Acta Optica Sinica **31**(12), 1212008 1-5 (2011).

[116] B. Edlén, "The refractive index of air," Metrologia **2**(2) 71-80 (1966).

[117] K. P. Birch and M. J. Downs, "An updated Edlén equation for the refractive index of air," Metrologia **30** 155-162 (1993).

[118] K. P. Birch and M. J. Downs, "Correction to the updated Edlén equation for the refractive index of air," Metrologia **31** 315-316 (1994).

[119] http://www.zygo.com/?/met/interferometers/dynafiz/

[120] D. M. Sykora, and P. de Groot, "Instantaneous interferometry: another view," OF&T Technical Digest, OMA1 (2010).

[121] D. M. Sykora, and M. L. Holmes, "Dynamic measurements using a Fizeau interferometer," Proc. SPIE **8082**, 80821R (2011).

[122] L. He, Y. Guo, H. Huang, Z. Wang, and J. Wang, "The technique of online measurement and data processing for aspheric machining," Advanced Materials Research **97-101**, 4313-4318 (2010).

[123] Y. Arai, W. Gao, H. Shimizu, S. Kiyono, T. Kuriyagawa, "On-machine measurement of aspherical surface profile," Nanotechnology and Precision Engineering **2**, 200-216 (2004).

APPENDIX: MATLAB CODE

```
% Calculate Two-quadrant Area Structure Function with sliding method
function [X,Y,ASF]=ASF(map)

close all;                  % Delete all figures whose handles are not hidden
clc;                        % Clear the display screen

figure
imagesc(map);               % Display input data as an image
xlabel('x')                 % Label the x-axis of the current axes
ylabel('y')                 % Label the y-axis of the current axes
title('Input Map')          % Add title to current axes
axis xy                     % Make the coordinate origin in the lower left corner
axis square                 % Make the current axes region square
colorbar                    % Show color scale

tic                         % Start a stopwatch timer

[dimv,dimh]=size(map);      % Size of map
map_flip=flipud(map);       % Flip input matrix up to down, for calculation in the second
                            % quadrant
Z1=nan(dimv,dimh);          % Preallocate matrix - set all Z1 to nan
Z2=Z1;

parfor i=0:dimh-1           % Parallel for-loop

    sf1=nan(dimv,1);        % Preallocate vector for Area SF in one column
    sf2=nan(dimv,1);

    for j=0:dimv-1
        differ=map(j+1:dimv,i+1:dimh)-map(1:dimv-j,1:dimh-i);   % Subtract map
                                                   % from a translated version of itself
        differ_flip=map_flip(j+1:dimv,i+1:dimh)-map_flip(1:dimv-j,1:dimh-i);
                                                   % Subtract map in the other direction

        % Calculate area SF in first quadrant
        k1=find(~isnan(differ));                   % Find indices of numerical elements
        diffout1=differ(k1);                       % Find numerical elements
        Z=sum(diffout1.^2)/length(diffout1);       % Area SF for i, j

        % Calculate area SF in second quadrant
        k2=find(~isnan(differ_flip));
        diffout2=differ_flip(k2);
        Z_flip=sum(diffout2.^2)/length(diffout2);
```

```
        sf1(j+1)=Z;
        sf2(j+1)=Z_flip;
    end

    Z1(:,i+1)=sf1;              % Area SF in one direction
    Z2(:,i+1)=sf2;              % Area SF in the other direction

end

SF1=Z1;                         % Area SF in the form of squared height difference
SF2=Z2;
% SF1=sqrt(Z1);                 % Area SF in the form of RMS
% SF2=sqrt(Z2);

SF_Right=SF1;                   % Area SF in the first quadrant
SF2_flip=fliplr(SF2);
SF_Left=SF2_flip(:,1:dimh-1);   % Area SF in the second quadrant


[X,Y]=meshgrid(1-dimh:dimh-1,0:dimv-1);
ASF=[SF_Left,SF_Right];         % Area SF in two quads

figure;
surf(X,Y,ASF);          % Draw area SF in two quadrants
xlabel('x(pixels)')
ylabel('y(pixels)')
zlabel('SF(\mum^2)')
title('Area SF (\mum^2) ')
shading interp;         % Vary the color in each line segment and face by interpolating
                        % the colormap index or true color value across the line or face
axis normal

toc                     % Obtain the elapsed time since tic was used
```

```
% Calculate Linear Structure Function with sliding method. (For non-circular aperture,
the deviation may increase.)

% Main function
function [r_final,SFr_final]=LSF(map)

close all;              % Delete all figures whose handles are not hidden
clc;                    % Clear the display screen

figure
imagesc(map);           % Display input data as an image
xlabel('x')             % Label the x-axis of the current axes
ylabel('y')             % Label the y-axis of the current axes
title('Input Map')      % Add title to current axes
axis xy                 % Make the coordinate origin in the lower left corner
axis square             % Make the current axes region square (or cubed when
                        % three-dimensional)
colorbar                % Show color scale

[r1,SFr1]=CalculationOneQuadrant(map);          % Use input matrix, call subfunction
[r2,SFr2]=CalculationOneQuadrant(flipud(map));  % Flip input matrix up to down, and
                                                % call subfunction

r=[r1;r2];
SFr=[SFr1;SFr2];

% Compute each element of "SFr_final" by averaging the "SFr" with the same "r"
[r_sort,IX]=sort(r);    % Return an array of indices IX, where size(IX) == size(r). As a
                        % vector, r_sort = r(IX).
SFr_sort=SFr(IX);       % SFr_sort is corresponding to r_sort

d=diff([r_sort;max(r_sort)+1]);
m=diff(find([1;d])) ;   % Obtain vectors m and n containing the frequency counts and
n=r_sort(find(d));      % the bin locations. It is faster than command "hist".

SFr_final=nan(length(n),1);         % Preallocation for "SFr_final" - set
                                    % each element to nan
r_final=nan(length(n),1);

ss(1)=m(1);
SFr_final(1)=sum(SFr_sort(1:m(1)))/m(1);
r_final(1)=n(1);
for i=2:length(n)
    ss(i)=ss(i-1)+m(i);
    SFr_final(i)=sum(SFr_sort(ss(i-1)+1:ss(i)))/m(i); % "SFr_final" obtained by
                                    % averaging "SFr_sort" with the same "r_sort"
```

```
        r_final(i)=n(i);                          % "r_final" corresponding to "SFr_final"
end

figure
plot(r_final,SFr_final)                           % Plot linear SF
xlabel('Separation');
ylabel('LSF')
title('Linear Structure Function');


% Subfunction
function [r_sub,SFr_sub]=CalculationOneQuadrant(map)

[dimv,dimh]=size(map);                            % Size of map
SFr=nan(dimv*dimh,1);                             % Preallocation for linear SF
r=nan(dimv*dimh,1);                               % Preallocation for the separation

p=0;

for i=0:dimh-1
    for j=0:dimv-1
        differ=map(j+1:dimv,i+1:dimh)-map(1:dimv-j,1:dimh-i); % Subtract map from a
                                                              % translated version of itself
        k=find(~isnan(differ));                   % Find indices of numerical elements
        diffout=differ(k);                        % Find numerical elements
        Z=sum(diffout.^2)/length(diffout);        % SF for i, j
        p=p+1;
        r(p)=sqrt((j)^2+(i)^2);                   % Point separation, but different indices
                                                  % of "r" may indicate the same separation
        SFr(p)=Z;                                 %"SFr" corresponding to "r"
    end
end

kk=find(~isnan(SFr));                             % Find numerical elements
r=r(kk);
SFr=SFr(kk);
r=round(r);

% Compute each element of "SFr_sub" by averaging the "SFr" with the same "r".
% The following part is similar as that in the main function. The computation speed is
increased when we put it in the subfunction.
[r_sort,IX]=sort(r);
SFr_sort=SFr(IX);

d=diff([r_sort;max(r_sort)+1]);
m=diff(find([1;d])) ;
```

```
n=r_sort(find(d));

SFr_sub=nan(length(n),1);
r_sub=nan(length(n),1);

ss(1)=m(1);
SFr_sub(1)=sum(SFr_sort(1:m(1)))/m(1);
r_sub(1)=n(1);

for i=2:length(n)
	ss(i)=ss(i-1)+m(i);
	SFr_sub(i)=sum(SFr_sort(ss(i-1)+1:ss(i)))/m(i);
	r_sub(i)=n(i);
end
```