IMPERIAL

Imperial College London

Department of Physics
Light Community

Optimising structured illumination microscopy

Meizhu Liang

Supervised by
Professor Mark A. A. Neil

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy (PhD)

June 2024

Abstract

Structured illumination microscopy (SIM) features high imaging speed, resolution beyond the diffraction limit and a large field of view, revolutionising conventional microscopy and is well-suited for long-term live-cell imaging and structural biology studies. The system presented in this thesis exploits hexagonal-patterned illumination using a spatial light modulator (SLM) to realise SIM.

The SIM imaging process is simulated and extended to cover various SIM geometries, where patterns of illumination are produced from polarised beams. All simulations are integrated as a plugin for the Napari platform and include sample movements such as defocusing, drift and Brownian motion.

In this SIM technique, the SLM provides the illumination beams by diffraction from displayed holograms. However, in practice, the diffracted orders were found to be extended due to the SLM not being perfectly flat. To measure and correct this aberration, a phase-recovery method is developed based on measurements of the diffracted PSF, which effectively corrects distorted diffracted orders to achieve diffraction-limited results. The resultant wavefront RMS 0.063 rad is smaller than the diffraction-limited value of 0.469 rad.

Three methods of hologram generation are compared. A fast method based on direct binarisation of beam interference (DBBI) is introduced. This method can generate holograms including phase corrections without compromising speed. Subtle features of the corresponding structured illumination can be further corrected by stepping an additional phase offset.

Our optical system is modified to allow remote and easy alignment. Phase recovery is performed in-situ to correct any changes in SLM flatness. The resultant PSF is diffraction-limited since the wavefront RMS is measured as 0.113 rad. Captured and reconstructed images with this correction show a complete field of view (FOV), while results without correction have a limited FOV. Additionally, phase stepping is applied to reduce patterned artefacts in the reconstructed images.

Declarations

Author declaration

I hereby declare that this thesis is my own work, with the exception of two figures reproduced from the sources given in the text, which are used with permission.

I acknowledge and give due credit to the works of others used or referenced in this thesis.
I further acknowledge the invaluable support and guidance provided by my supervisor Prof. Mark A. A. Neil.

Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Contributions

The HexSIMInstrument characterised in Chapter 5 was based on an earlier version (including the framework of controlling the Hamamatsu camera, lasers and the measurement window) originally written by Hai Gong. Previously developed code (by Hai) was used to control lasers. I changed and refined the GUI designed by Hai to control the DAQ box, the SLM, the Hamamatsu and the Thorlabs cameras. In terms of my experiments, the optical set-up was redesigned and implemented. In addition, I installed and operated all phase-recovery equipment and wrote all controlling code. I also developed all phase recovery algorithms and the DBBI methods.

Conference presentations

- M. Liang, Prof. Mark A. A. Neil, “Phase recovery for measuring and correcting Spatial Light Modulator (SLM) aberrations in SIM microscopy” Talk presented at Focus on Microscopy 2024
- M. Liang, Prof. Mark A. A. Neil, “Structured illumination microscopy” Invited talk in Control, Optimisation & Learning Seminar Series, Zhejiang University, 2024.
- M. Liang, Prof. Mark A. A. Neil, “A custom plugin: napari-generic-SIMulator” Invited talk in napari workshop, Politecnico di Milano, 2023.

Software published

- napari-generic-SIMulator

  https://www.napari-hub.org/plugins/napari-generic-SIMulator

To my mum and dad

献给我的妈妈和爸爸

Acknowledgements

Completing this six-chapter thesis has been a significant chapter in my life, and it would not have been achievable without the guidance and support of many individuals to whom I am deeply grateful.

I would like to thank my supervisor Prof. Mark A. A. Neil from the bottom of my heart for accepting me as his PhD student. Your wisdom, guidance and support have not only shaped my research but also profoundly influenced my personal growth. I am grateful for your openness in allowing me to 'barge into' your office whenever I needed help.

I would also thank Andrea for the guidance in working on designing napari plugins, to Chris for giving me helpful suggestions and being my assessor at each milestone of my PhD, to Martin and Simon for customising optical hardware, to Nigel Cartwright and Henning Molsen from Forth Dimension Displays for helping with updating SLM's functions, to Hai and Wenjun for helping and guiding me in my early PhD life, to Judith, Marcia and Loli for helpful guidance in admin work, to Zhonghe for explaining physics to me, helping with the equipment and preparing lab documents, to Yuheng for creating a great place to work in our lab.

I would also like to extend my thanks to Sunil and Ara for drinks. I've lost count of how many glasses I owe you two. Also, to Huihui for accompanying me to taste the bittersweet PhD life (and takeaways), to Wenzhi for the encouragement and support when I face challenges, to Bill, Jonny, Leo, Liuba, Anita, Jack, Yuchen and Cheryl for stress-relieving chats.

I would additionally like to express my gratitude to all the present and past members of the Photonics group: Paul, James, Carl, Mike, Martin, Edwin, Hugh, Dan, Shengjie, Yuiry, Sara, Karishma, Matt, Huaifeng, Xunuo, Abdual, Tugba, Yu, Bartosz, Kenny, Nathan, Niall, Nils, Miqi, Wim, Goronwy, Will.

To my beloved Mum and Dad, your encouragement during moments of difficulties and your financial support have been instrumental in my academic success. To dear Meizhen, thank you for accompanying me not only through my doctoral years but also through all the years until now. To dear Connor, thank you for having good taste and being supportive. To all my family and Jane, Alan, Lucy and Hal, your love and support have been the guiding light throughout my PhD journey.

Contents

Chapter 1. Thesis overview 21

Chapter 2. An introduction to structured illumination microscopy 25

2.1. From epifluorescence microscopy to confocal microscopy 25

2.2. Structured illumination microscopy 29

2.3. Optimisation of an existing SLM HexSIM system 38

2.4. Increasing resolution further in optical microscopy 40

2.5. Chapter summary 41

Chapter 3. Theory and simulation of SIM imaging 43

3.1. Simulating image formation in a SIM microscope 43

3.2. Calculating the illumination intensity in the sample space 45

3.3. imaging from a sample space to a camera plane 50

3.4. Vectorial calculation to take account of dipole emitters 56

3.5. Computing the simulation of SIM imaging 58

3.6. Implementation as a napari plugin: napari-generic-SIMulator 64

3.7. napari-generic-SIMulator Results 69

3.8. Chapter summary 80

3.9. Appendix: Nyquist sampling in the pupil plane and the PSF plane 81

Chapter 4. Correction for the flatness of R11-2K SLM 85

4.1. The distorted diffraction order 85

4.2. Phase recovery for measuring and correcting SLM aberrations 87

4.3. Chebyshev polynomial expansion for a square aperture 98

4.4. Simulation results 109

4.5. Implementation with a simple optical layout 117

4.6. Chapter summary 124

Chapter 5. Hologram for HexSIM microscopy and aberration correction 127

5.1. Using an SLM to generate SIM patterns 127

5.2. Generating holograms for 3-beam HexSIM illumination patterns 131
5.3. Diffraction efficiency and the pixel balance problem 142
5.4. Uniformity and coverage of illumination at the sample plane 151
5.5. Applying phase stepping in the DBBI method 155
5.6. Pinhole size 163
5.7. Chapter summary 169
Chapter 6. Implementation and measurements 173
6.1. Optical system and platforms 173
6.2. Implementations in HexSIM ScopeFoundry 179
6.3. Results of in-situ phase recovery in the HexSIM system 184
6.4. Measurements on fluorescent samples 189
6.5. Chapter summary 200
Chapter 7. Conclusion and future work 202
7.1. Conclusion 202
7.2. Future work 204
References 208

List of Figures

Figure 2-1. (a) Schematic of a typical epifluorescence microscopy system. (b) Schematic of a confocal microscopy system. Relay lenses are emitted in images. 26

Figure 2-2. The OTF in (a) conventional microscopy and (b) confocal microscopy. The centre of the OTF is denoted using a circle. 28

Figure 2-3. Schematic of the optical layout. (b) Axial response measurements in the system. Reprinted with permission from [21] © Optical Society of America. 30

Figure 2-4. The Moiré effect in daily life. (a) demonstrates the pattern of small round wholes at the front and the back of a chair. (b) The overlapped patterns of both sides of the chair's back are Moiré stripes, which are denoted by arrows. Coarse stripes can also be observed from yellow chairs. 32

Figure 2-5. (a) The field amplitude of two beams. (b) The intensity of the Fourier transform of the two beams. (c) The interference between two beams. Distances between components and the centre are denoted using arrows. All complex fields are shown only with real amplitude values in the figure. 34

Figure 2-6. Principle behind resolution enhancement in conventional SIM microscopy. (a) The sample is superposed with the structured illumination pattern, where frequencies are denoted. (b) The detectable area of the conventional microscope system in reciprocal space, which is can also be seen as the OTF support region. (c) Two frequency components of the illumination pattern are added into the observable area and denoted in green and yellow. The zero-frequency component located at the centre is shown as a black spot. (d) The two frequencies extend two offset regions based in reciprocal space. (e) The detectable region is expanded to a twice size by rotating the structured illumination pattern. 34

Figure 2-7. The 3D OTF in (a) conventional microscopy and (b) a two-beam SIM microscope. The shifted OTFs are in grey. The centre of the OTF is denoted using a circle. 36

Figure 2-8. (a) The field of three beams. (b) The intensity of the Fourier transform of the three beams. (c) The inverse Fourier transform of the interference intensity. Distances between components and the centre are denoted using arrows. All complex fields are shown as amplitude values in the figure. 36

Figure 2-9. The resolution improvement in HexSIM microscopy. (a) A hexagonal mask for generating a hexagonal illumination pattern. (b) The detectable area of the conventional microscope system in reciprocal space and frequency components of the hexagonal illumination. The zero-frequency is denoted as a black spot, while the other three are shown in blue, green and yellow. (c) The detectable region is enlarged by applying the structured

illumination pattern without rotations. Conjugates of frequency components and the DC component are omitted in the figure. 37

Figure 2-10. The set-up of the previous HexSIM system based on FLCSLM. Figure is adopted from the supplementary document of [38] with permission. 39

Figure 3-1. Schematic of a SIM system. The light path of the illumination and the emission is demonstrated in green and yellow, respectively. 44

Figure 3-2.Beams in different SIM modes. The grey circle denotes the illumination pupil. 46

Figure 3-3. Polarisation states. The number of beams is taken as 4, which can be adjusted in the SIMulator. 48

Figure 3-4. Four examples of illuminations with different polarised beams. The illumination changes with different parameters and situations in the SIMulator. 50

Figure 3-5. The fields close to the front and back focal plane of a lens. 51

Figure 3-6. The layout of the microscope for imaging from a sample space to a camera plane. 52

Figure 3-7. (a) A dipole radiation at a random direction. (b) Dipole radiation at x-y plane and x-z plane (c). the radiation changes with the observation direction of the objective in a system. 56

Figure 3-8. A slice of 3-D samples: Spheroidal (a) and filament (b) point-clouds. 59

Figure 3-9. Dodecahedron and directions of single emitters (blue dots). The distribution Over the dodecahedron is isotropic. The left dodecahedron is randomly rotated compared to the right one. 62

Figure 3-10. GUI of napari-generic-SIMulator. (a) Point-cloud generator widget in a napari viewing window. (b) SIM data generator widget. The two widgets are usually tabified. 65

Figure 3-11. Generation of a SIM raw stack in SIMulator. Steps with the blue rotating arrow need to be repeated. 66

Figure 3-12. Two examples of simulated images: a spheroidal point-cloud in x-y plane (a) and x-z plane (c); a filament-point-cloud in x-y (b) and x-z plane (d). 67

Figure 3-13. The superposition of a spheroidal sample point-cloud and the corresponding raw image in x-y (a) and x-z (c) plane. Filaments overlapped with the point-cloud in x-y (b) and x-z plane (d). Point-cloud samples are shown in green. 68

Figure 3-14. The slice of a scalar PSF in $x - y$ plane (a) and $y - z$ plane (b). 70

Figure 3-15. Vectorial PSFs: The PSF stack of flexible dipoles in $x-y$ (a) and $y-z$ plane (b). The PSF calculated with rigid dipoles in the horizontal (c) and vertical plane (d). Apart from the PSF calculation, all parameters are the same as those used to simulate the scalar PSF. 70

Figure 3-16. Scalar OTF in $kx-ky$ (a) and $ky-kz$ planes (d). The horizontal (b) and vertical (e) slice of the flexible OTF. There is an anulus of zero in the rigid OTF with z-polarised excitation that can be seen slices in the two planes presented in (c) and (f). 71

Figure 3-17. Parameters of the simulated raw image stack in (a) point cloud generator and in (b) SIM data generator. 72

Figure 3-18. napari-SIM-processor GUI. The loaded image stack is the filament-point-cloud simulated above. Parameters shownin the figure are default values, which need to be changed according to the raw images. 73

Figure 3-19. Raw images along z axis with 1st phase and 1st orientation. Top row: 1st -5th frame. Middle: 6th -10th frame. Bottom: 11th -15th frame. The distance between displayed frames is 50nm. 74

Figure 3-20. Raw frames with different phases and the differences. Top row: 3 raw images with 3 phases: P1, P2 and P3 at the first angle: A1. Bottom: differences between frames. ..74

Figure 3-21. Three raw images at different angles (top) and the frame-differences (bottom) at the 2nd phase (P2). 75

Figure 3-22. Cross-correlation of the raw image that is at the middle of the z-stack. The red circle in (a) (b) and (c) denotes the location of the frequency order at the 1st, 2nd and 3rd angle, respectively. 76

Figure 3-23. Two calibration results provided by the Napari-SIM-processor: the raw image spectrum (a) and the Wiener filter used in reconstruction (b) of the enhanced resolution image. 77

Figure 3-24. (a) The wide-field image. Reconstruction of the simulated raw image using conventional SIM (b) and batchSIM (c). (d) Intensity profile along yellow lines in (a), (b) and (c). 78

Figure 3-25. Raw images and reconstructed frames of different SIM geometries with various beam polarisation states. Left column: a raw image. Right column: super-resolved images. 79

Figure 4-1. Optical layout for measurement of diffraction orders. A low-power laser is used as the light source. 86

Figure 4-2. (a) The image of a diffraction order. The pinhole is denoted as a white ring. (b) The installation of the SLM. Two screws on the mounting frame are emphasised using blue circles. 86

Figure 4-3. Gerchberg-Saxton algorithm. Red annotations are related to three introduced wavefront biases. 88

Figure 4-4. Preparation of the phase retrieval with Chebyshev polynomials. (a) Target pupil phase φ. (b) Wrapped pupil phase Φ. (c) Initial guess of the unwrapped pupil phase φg. (d) Wrapped estimated phase $\Phi \mathrm{g}$. The pixel number of each phase is $N(256) \times N(256)$. The unit of the colour bars is radians. 90

Figure 4-5. Three wavefront biases and corresponding PSFs. (a) The positive defocus $b0$. (b) The positive astigmatism $b1$. (c) The negative defocus $b2$. These are all central symetric. (d) – (f): Corresponding PSFs. 92

Figure 4-6. Simulated measurement of three initial PSFs corresponding to the same pupil phase with distinct added biases of positive defocus (a), astigmatism (b) and minus defocus (c). 93

Figure 4-7. Estimating the pupil plane based on PSFs by the G-S algorithm. (a) Target phase φ. (b) Unwrapped phase estimate φGS. (c) The target PSF. (d) Estimated PSF. All results are acquired after the first 75 G-S iterations. 95

Figure 4-8. Phase smoothing. (a) Unwrapped phase (φGS) retireved by the first G-S iteration. (b) Unwrapped phase fitted by Chebyshev polynomials: φc. 96

Figure 4-9. Flowchart of the phase recovery. 97

Figure 4-10. Chebyshev polynomials of the first kind. 99

Figure 4-11. 2-D Chebyshev polynomials Bij sampling at nodal points. The subscripts denote the location of the polynomial values in the matrix plot. 102

Figure 4-12. Matrix plots of the orthogonality relation and the normalisation of Chebyshev polynomials approximating the phase and the amplitude of the pupil plane. (a) The discrete orthogonality of the phase ones. (b) Sums of the multiplying products of 2 polynomials and the normalisation values. The orthogonality matrix of Chebyshev polynomials for fitting the pupil amplitude without (c) and with the normalised coefficients (d). 104

Figure 4-13. The weighting values over the pupil plane. The four vertices of the square are endpoints of the mask. 107

Figure 4-14. Two-dimensional Chebyshev polynomials $Bwij$ over the pupil plane with a 2-D mask. 108

Figure 4-15. The orthogonal relation of Chebyshev polynomials obtained with weighted summation. The normalised orthogonality of the Chebyshev expressions is shown both before (a) and after (b) correction of the weighting factor at the end points. 109

Figure 4-16. Phase recovery with sampled Chebyshev fitting. (a) The calculated PSF with the aberration and 0th bias: $psfGS0$. (b) The estimated G-S pupil phase: φGS. (c) φGS minus the actual phase aberration Φ. (d) The estimated pupil amplitude. (e) The rms of the current fitted phase φc. (f) The wrapped simulated phase aberration: Φ. (g) Unwrapped φGS. (h) The chebyshev fitted phase: φc. (i) The wrapped accumulated phase. (j – l) The simulated measurments of corrected PSFs with the 0th, 1st and 2nd bias, respectively. Phases in (c), (f) and (i) are wrapped to cover a range from 0 to 2π by taking the angle function to present more straightforward images. 110

Figure 4-17. The progression of a typical phase recovery. The row from top to bottom shows the calculated PSFs, simulated PSF measurements, estimated phases, unwrapped phase estimations, Chebyshev-fitted phases, accumulated phases, and the pupil intensities, respectively. 111

Figure 4-18. Phase correciton with the weighted Chebyshev approach. (a) The calculated PSF with the aberration and 0th bias. (b) The estimated G-S pupil phase. (c) φGS minus the actual phase aberration Φ. (d) The pupil amplitude estimate. (e) The rms of the current fitted phase φc. (f) The wrapped simulated phase aberration: Φ. (g) Unwrapped φGS. (h) The chebyshev fitted phase: φc. (i) The wrapped accumulated phase. (j – l) The simulated measurments of corrected PSFs with the 0th, 1st and 2nd bias, respectively. Phases in (c), (f) and (i) are wrapped to cover from 0 to 2π by taking the angle function to present more straightforward images. 112

Figure 4-19. PSFs before (a) and after the phase recovery with the sampled (b) and the weighted (c) Chebyshev fittings. 113

Figure 4-20. Simulated parameters in the phase recovery with the Zernike fitting. (a) φ: The simulated phase aberration at the SLM plane. (b) $b0$: $7.5 \times Z20$. (c) $b1$: 0. (d) $b2$: $-7.5 \times Z20$. (e) φg: The initial guess of the phase aberration. (f) - (h): The PSF corresponding to the aberrated wavefront with biases $b0$ - $b2$. 114

Figure 4-21. Phase recovery with Zernike fitting. (a) The calculated PSF with the aberration and 0th bias: $psfGS0$. (b) The estimated G-S pupil phase. (c) φGS minus the actual phase aberration Φ. (d) The estimated pupil amplitude. (e) The rms of the current fitted phase φc. (f) The wrapped simulated phase aberration: Φ. (g) Unwrapped φGS. (h) The chebyshev fitted phase: φc. (i) The wrapped accumulated phase. (j – l) The simulated measurments of

corrected PSFs with the 0th, 1st and 2nd bias, respectively. Phases in (c), (f) and (i) are wrapped to present more straightforward images. 116

Figure 4-22. Flowchat of the hardware triggering in capturing diffracted orders. All devices are controlled by the Python code by using the corresponding packages. Green arrows denote hardware signals between devices. 119

Figure 4-23. Hardware triggering signal of devices. The start of each hologram display is synchronised with the exposure of the camera 120

Figure 4-24. The diffraction orders with applied three biases. 121

Figure 4-25. Measuring and recovering the phase of the SLM's surface with the simple layout. (a) – (c) Three PSFs with introduced biases. (d) The calculated PSF with the 0th bias. (e) The calculated pupil phase with rms of 0.412 radians. (f) The calculated pupil intensity. (g) The matrix of values of Chebyshev modes in the current step, with the square on the top left showing the piston value. (h) The estimated unwrapped phase with RMS of 0.412 radians. (i) The Chebyshev fitted phase with RMS of 0.063 radians after excluding piston and tilt. (j) Progression of the rms of the fitted pupil phase over the correction process. (k) The total estimated phase accumulation. (l) The hologram written into the flash of the SLM to produce a corrected diffraction order as shown in (a). All results are obtained from the last correction iteration. 123

Figure 4-26. Comparison between the initial (a) and the corrected (b) measuremtns of the diffraction order. 124

Figure 5-1. Frequencies in the Fourier domain. Three field components (blue) and three intensity components (orange). Conjugate components are omitted from the image. 130

Figure 5-2. Polka dots. Periods of the pattern in x and y directions are annotated. 132

Figure 5-3. Parameters of dots pattern. (a) The relationship between the pitch and the radius of a polka dot. p and $r0$ is annotated in green and yellow, respectively. The image is cropped from Figure 4-21 and is zoomed in. (b) The shift in seven patterns is displayed by blue arrows. The dark and light green star denotes the original and the final position of a polka dot, respectively. The positive and negative areas are opposite to (a), but the pattern is still binary. 134

Figure 5-4. Zoomed-in cropped seven holograms produced by the polka dots approximation. The holograms are rotated with $\pi 10$ rad to show the sptial movement more clearly. 135

Figure 5-5. Schematic of a lens performing Fourier transform. A hologram loaded on the SLM can be seen as a grating with the period as d. 136

Figure 5-6. (a) A hologram diffracting the light into three directions denoted in different colours. (b) The schematic of the corresponding diffracted orders in the mask (the grey disc) plane. Only zeroth and the first orders are shown. The size of the mask, pinholes and the distance between beams in the image are arbitrarily chosen here to illustrate the layout. The pink arrow in (a) and (b) annotates the distance in the SLM plane and the mask plane, respectively. . 136

Figure 5-7. Zoomed-in cropped holograms with different pitch p in pixel. (a) p = 10. (b) $p = 8.229$. This is a value calculated with an actual optical layout. $\lambda = 561\ nm$, $f = 250\ mm$, $dm = 2.4\ mm$ and $d = 8.2\ \mu m$ are substituted in equation 5-16 to calculate the p. (c) p = 6. 137

Figure 5-8. Zoomed-in cropped seven holograms produced by the binarisation of the interference between three beams. The bemas phase φbh changes with the number of hologram while both φc and φcon keep as zero. 140

Figure 5-9. Zoomed-in cropped holograms with different constant phase φcon. (a) $\varphi in = 0$. (b) $\varphi in = \pi 6$. (c) $\varphi in = 2\pi 6$. (d) $\varphi in = 3\pi 6$. The correction phase φc for all holograms is zero. 140

Figure 5-10. The hologram with the phase level as 2 (a), 4 (b) and 8 (c), respectively. All images are obtained by taking the real part of the field retrieved by the binary research. 141

Figure 5-11. Diffraction orders and the correspondent illumination. (a) shows the diffraction patterns passing through the three pinholes. (b) shows the zoomed in pinhole which locates on the right, and (c) displays the illumination from the pinhole. As the hologram is rotated with $\pi 10$, the three orders shown in (d). (e) and (f) show the zoomed in beam and the illumination after the rotation, respectively. All results are calculated with holograms generated by the polka dots approximation. 143

Figure 5-12. Results produced using three hologram generation methods. The hologram is generated by (a) polka dots approximation, (b) DBBI method and (c) the optimisation of the binary search, respectively. (d), (e) and (f) show the square root values of the diffracted field to highlight diffraction orders. The absolute values of the field of diffraction beams are shown and zoomed in further in (g), (h) and (i). All images are cropped and enlarged to show the patterns more clearly. Red circles denote the positions of pinholes on the mask..... 144

Figure 5-13. Comparison of the diffraction intensity between the polka dots method, DBBI and the binary search optimisation. 146

Figure 5-14. Comparison between the simulated aberration and the correction. (a) The aberration in the SIM system. (b) The retrieved correction. (c) The difference between the correction and the aberration. The unit of colour bars is radians. 147

Figure 5-15. (a) The hologram generated by the polka dots approximation with a 4:6 ratio. (b) The hologram generated by DBBI with the phase corection. The yellow area acounts for 50.12% of the whole area. (c) The hologram generated by the optimisation of binary search. 148

Figure 5-16. The image of the light diffracted by the hologram generated by the 4:6 polka dots method (a), by DBBI including the phase correction (b) and the binary search approach (c). Red circles show locations of the pinholes on the mask. 149

Figure 5-17. Comparison of the diffraction intensity between the 5:5 polka dots method, DBBI, the optimisation and the 4:6 polka dots approach. 149

Figure 5-18. The intensity of the light going through one pinhole in the system with the aberration. The light diffracts at the hologram produced using the 4:6 polka dots approximation (a) and DBBI method (b). The hologram in (b) is generated with the constant phase to be zero. 152

Figure 5-19. The interference between diffraction orders passing through three pinholes when using a hologram generated by the DBBI method. (a) The three pinholes are shown as yellow circles. The number of each pinhole is denoted in white. (b) The DC magnitude generated by the hologram. (c) The magnitude interference between the pinhole 1 and 2 ($m1m2$). (d) $m2m3$. (e) $m3m1$. The unit of colour bars is radians. 154

Figure 5-20. (a) The magnitude of the interference between the pinhole 1 and 2 ($m1m2$). (b) $m2m3$(c) $m3m1$. The hologram is produced using 4:6 dots approximation. The unit of colour bars is radians. 154

Figure 5-21. The interference between the light transmitting through two pinholes with the misalignment. (a) to (c): The interference between the 1st and the 2nd, the 2nd and the 3rd, and the 3rd and the 1st carriers, respectively, with the hologram generated with the 4:6 dots approximation. (d) to (f): $m1m2$, $m2m3$ and $m3m1$in DBBI case. 155

Figure 5-22. The summed illumination through a single pinhole generated using the phase stepping applied in the binarisation of interference between beams with one step (a), 10 steps (b) and 50 steps (c). 156

Figure 5-23. The 2nd illumination pattern generated by the interference between diffraction orders passing through the three pinholes. (a) The second structured light pattern. (b) The cropped and zoomed-in part of the 2nd illumination pattern. (c) The 2nd carrier field obtained by the Fourier transform of seven illumination intensities along the axis of seven points. 157

Figure 5-24. (a) The frequency values of the second carrier field. The carrier 2 is dominant here, so it is much brighter than the others. Seven carriers in the frequency domain can be

seen and are denotated in white. Log function is used to emphasise the carriers in the image, and the image is cropped. (b) The zoomed-in 2nd frequency carrier. 158

Figure 5-25. Amplitudes and phases of the frequency carriers. (a) to (c): The amplitude of the first, second and third carrier divided by the zeroth frequency, respectively. (d) The intensity of the 0th carrier. (e) The 10x zoomed-in cropped phase of the second carrier (from the centre of (h)). (f) The 10x zoomed-in cropped of the amplitude of the third carrier (see (c)). (g) to (i): The residual phase of the 1st, 2nd and 3rd frequency components, after the expected tilt is removed. 159

Figure 5-26. The standard deviation σ of the amplitudes and phases of three frequency components with phase steps in an anerration-free situation. The range of the phase steps is from one to twenty. The holograms are produced using the DBBI method. 160

Figure 5-27. The standard deviation σ of the relative amplitudes and residual phases of three carriers with 20 phase steps with the aberration obtained from the optical setup. The method to generate holograms is taken as the DBBI with the phase correction. 162

Figure 5-28. The amplitude of the light orders diffracted by holograms generated by the 4:6 dots approximation (top row) and the DBBI method (bottom row) with different sizes of pinholes. The amplitude of the diffraction orders with utilising a pinhole with the radius as (a) $100\ \mu m$ (which is the same parameter in our practical mask), (b) 200 μm and (c) $50\mu m$. (d) – (f) are diffracted orders of a hologram generated by the DBBI approach with the constant phase to be zero, passing through pinholes with different radii. 164

Figure 5-29. The interference between diffraction passing through (a) the 1st and the 2nd, (b) the 2nd and the 3rd, and (c) the 3rd and 1st with a pinhole radius of $200\ \mu m$ and the hologram generation as the 4:6 polka dots approximation. (d) to (f): $IN12$, $IN23$ and $IN31$, respectively, with the $200\ \mu m$ radius but exploiting DBBI. 165

Figure 5-30. The interference between diffraction passing through two pinholes with the radius as as $50\ \mu m$ and the hologram design as DBBI. (a) $IN12$. (b) $IN23$. (c) $IN31$. The holograms are generated by DBBI. 165

Figure 5-31. Frequency carriers with $200\ \mu m$ pinholes. Holograms are produced by the 4:6 polka dots approximation. (a) The frequency values of the 3rd carrier field. Seven carriers in the frequency domain can be seen and the 3rd (denoted in the orange frame) is dominant. Log function is used to emphasise the carriers in the image. (b) The zoomed-in 3rd frequency carrier. (c) The zoomed-in 1st frequency carrier, which is framed in a green rectangular. ... 166

Figure 5-32. Variations in the relative phases and residual phases with 20 phase steps when the radius is (a) 200 μm and (b) 50 μm. The method to generate holograms is taken as the DBBI with the phase correction. 168

Figure 6-1. Schematic of the HexSIM system. Only main equipment is shown in the image, some parts such as relay lenses, the microscope frame and some supporting equipment (e.g. a data acquisition box) are omitted. Seven holograms with phase correction are presented. 174

Figure 6-2. Aligning the diffracted orders with the pinhole mask. The image is captured by the XIMEA camera as an LED is illuminating the mask plane. In this image, the current positions of the diffracted beams require adjustment to align with the pinholes. 176

Figure 6-3. The block diagram of all the different device driver elements in our HexSIM system. 177

Figure 6-4. The graphical user interface (GUI) of HexSIM ScopeFoundry, including pieces of hardware, measurements, windows, functional buttons and so on. All tabs are collapsed in the screenshot, but can be expanded to control particular devices or measurements. 178

Figure 6-5. The control tab of the Thorlabs camera in the GUI of HexSIM ScopeFoundry. 178

Figure 6-6. The block diagram of the configuration of synchronising different device for SIM imaging. The initial step is written in a light grey shade. 180

Figure 6-7. The synchronisation between the SLM and the Hamamatsu camera. The holograms for each frame at a channel (such as the Green 0 and the Red 0) are different, but they are shown using the same symbol of coloured lines. The light lilac areas demonstrate the exposure of the camera from the first line to the line at the bottom of the frame. The synchronised imaging for one frame of a raw image in two channels is presented in the figure, respectively. Displaying inverse holograms for the maintenance of the SLM's DC balance is omitted in the figure for clarity. 182

Figure 6-8. The progression of the in-situ measurement and the recovery the phase of the SLM on the HexSIM layout. The phase recovery process is depicted from the left column to the right. 186

Figure 6-9. The in-situ measurement and the recovery the phase of the SLM on the HexSIM layout. (a) – (c) Three corrected PSFs with introduced biases. (d) The calculated PSF with the 1st bias. (e) The calculated pupil phase with rms as 0.340 radians. (f) The calculated pupil intensity. (g) The padding of values of Chebyshev modes, and the square on the top left shows the piston value. (h) The estimated unwrapped phase with rms as 0.340 rad. (i) The Chebyshev fitted phase with the 0.113 rms. (j) The rms of the fitted pupil phase. (k) The estimated phase

accumulation. (l) The hologram written into the flash of the SLM to diffract the order in (a). All results shown are obtained from the last correction iteration. 187

Figure 6-10. The diffraction beams locating at the mask before (a) and after (b) applying phase recovery to correct for SLM flatness. The two white rectanguls are used to denote the diffracion orders passing through the pinholes in two cases. Images are taken with camera 3. 187

Figure 6-11. Diffraction orders captured by the Thorlabs camera. (a) Uncorrected diffraction orders. (b) Corrected diffraction beams. The extended and weak three diffraction orders are conjugate to thoses are corrected. Images are over exposed to display weak orders. 188

Figure 6-12. (a) One beam without the phase correction. (b) One recovered diffraction order. The diameter of the pinhole on the mask is 200 μm. 189

Figure 6-13. Calibrated phase versus hologram step. Green channel (a), (b) and (c) and red channel (d), (e) and (f). Without correction (a) and (d), with correction (b) and (e), with correction and phase stepping (c) and (f). DBBI method used for all hologram designs. The RMS error is shown in each sub-image. 190

Figure 6-14. Wide-field images for four different cases of hologram generation with different modulation strategies. (a) The wide-field image with holograms generated by the DBBI method. (b) The DBBI with phase stepping is utilised to generate holograms. (c) Holograms are produced using the DBBi method with phase correction. (d) The wide-field image with the phase correction and phase stepping. All holograms used to construct illumination patterns are generated using the the DBBI approach. The four wide-field images are produced using napari-sim-processor as the zero-order carrier image of the acquired stack. 191

Figure 6-15. The spectrum of reconstructed images in green channel without phase correction(a), with phase correction (b) and with both phase correction and phase stepping (c). (d), (e) and (f) shows the corresponding spectrum of the reconstructed images from the red channel. 193

Figure 6-16. Reconstructed images using the DBBI method in different situations. (a) The case using DBBI without any modulations. (b) Utalising DBBI with the phase recovery. (c) The DBBI method is exploited with phase correction and phase stepping. (a1) – (c1): The yellow frame in (a) – (c), respectively. All imgaes are reconstructed using napari-SIM-processor. 194

Figure 6-17. Comparison of reconstructured HexSIM images. The holograms are produced using the DBBI method without (left column) and with phase stepping (right column), where the phase aberration is corrected. (b2) and (c2): The red-channel information with and without phase stepping, respectively. (b3) and (c3): The green-channel image exluding phase steeping

and including phase stepping, respectively. The blue and pink frames correspond to those in Figure 6-16 (b) and (c). 195

Figure 6-18. Wide-field images (a), (c), (e) and reconstructed images (b), (d), (f) of 100 nm beads in the green channel. (The excitation laser features 488 nm. The distance between a pinhole to the centre of the spatial mask is 2.3 mm.) The first row shows images without any corrections, the second row shows results with phase recovery, and the third row shows results with both phase recovery and phase stepping. All holograms are generated using the DBBI method. 197

Figure 6-19. Wide-field images (a), (c), (e) and reconstructed images (b), (d), (f) of 100 nm beads in the red channel (The excitation laser features 561 nm). The first row shows images without any corrections, the second row shows results with phase recovery, and the third row shows results with both phase recovery and phase stepping. All holograms are generated using the DBBI method. 199

Figure 6-21. Residual results in two channels. The first, second, and third elements in the sub-caption are the imaging channel, the phase situation, and the residual RMS, respectively. 200

Figure 7-1. (a) Schematic of diffracted orders with hologram beam shifting. A single beam (denoted by a green star) is diffracted into each pinhole by each hologram in a sequence that dithers the spot positions. (b) Schematic when applying the DBBI with correction in a conventional SIM geometry with sinusoidal pattern illumination. The centre of the diffracted orders (the optic axis) is denoted in green and that of the mask in black. 205

Chapter 1. Thesis overview

The topic of this thesis is the development and optimisation of a structured illumination microscopy system with hexagonal illumination and a ferroelectric liquid crystal spatial light modulator (FLCSLM HexSIM).

Chapter 2 provides a background to the work, by giving a theoretical overview of SIM microscopy and placing it in the broader context of fluorescence microscopy techniques. The chapter starts by introducing epifluorescence microscopy, as it is the basis of many microscopy techniques and so can be regarded as conventional microscopy. Confocal microscopy is then presented, which is also used to image fluorescence-labelled samples. This technique uses a pinhole to surpass the (Abbe) resolution limit in conventional microscopy and allows optical sectioning. Then, the concept of SIM microscopy is presented, which was originally utilised for optical sectioning. This thesis, however, focuses on super-resolution SIM microscopy, which improves the lateral resolution beyond the Abbe resolution limit. The principles behind the resolution enhancement in SIM using sinusoidal and hexagonal illumination patterns are explained, respectively. An existing SIM system is then described, which provides a baseline for the optimisations presented in this thesis. Although SIM microscopy can break the diffraction limit, some techniques (e.g., nonlinear SIM, STORM and STED) can outperform it in terms of resolution improvement, and these are briefly discussed.

Both the in-depth theory and simulations of imaging in a SIM microscope are presented in Chapter 3. This theory is leveraged in the development of a Python-based plugin, napari-generic-SIMulator, which is used to simulate SIM imaging and generate raw images. The development of this plugin is discussed in detail in this chapter. This tool is particularly useful for investigating SIM imaging involving sample movement and Brownian motion, as these can be simulated in the raw images, and the robustness of e.g. a reconstruction algorithm can be assessed. After briefly introducing the theory and simulation of image formation in a SIM microscope, the structured illumination is calculated. The different polarisation states of illuminating beams (horizontal, vertical, azimuthal, radial and circular) rotated by the objective lens in a SIM system are also calculated. The polarised beams are located at different points in the pupil of the objective lens, according to the particular one of the various SIM geometries (conventional 2-D, hexagonal, right-angle hexagonal and 3-D SIM) selected. The fluorophores in the sample are illuminated by the structured light and thus emit fluorescent light. This response is calculated, which starts with deducing the field in the pupil plane for a single point

in the sample. Then, the image of the point (the intensity in the camera plane) is calculated based on the pupil field, which is then extended to obtain the image of a point cloud. To cope with more complex cases, consideration of the fluorophores as orientated dipole emitters is also taken into account. Raw SIM images are computed, including simulating point clouds, modelling dipole orientations in a dodecahedral arrangement and calculating scalar, rigid and flexible PSFs and OTFs. The simulations with various parameters in different scenarios are realised using napari-generic-SIMulator, and the protracted phase-tilt calculation that is the time limiting stage in the calculation, can be accelerated either in CPU or GPU. The simulated raw images are shown and then processed by another bespoke plugin: napari-SIM-processor. Based on the algorithms in the processor, the SIM raw frames are reconstructed to be super-resolved, and the artefacts caused by sample movements are examined.

Chapter 4 presents work motivated by a practical problem in the current SIM system: the SLM is not flat enough, which causes diffraction orders to be extended and not diffraction limited. To address this issue, a phase recovery method is proposed to measure and correct the resultant phase aberration, which combines a Gerchberg-Saxton algorithm and PSF measurements. However, the aberration is usually too large for a solution to converge, and there are phase singularities in the retrieved results. Hence, a fitting function is utilised to smooth the phase, and more PSF measurements are taken as feedback via the optical system. For a square aperture, sampled and weighted Chebyshev polynomials are calculated and can be used for smoothing the phase. This phase correction approach is implemented both in simulations and in a practical optical layout. In either scenario, the diffracted orders can be recovered to a diffraction limited result. Details of the experimental implementation, such as writing holograms to the SLM's flash memory on the fly and hardware triggering, are explained.

In Chapter 5, a mini review is given to introduce SIM systems that use an SLM, and the principle of a ferroelectric liquid crystal SLM is briefly explained. The three-beam HexSIM illumination, which is used in our system, is introduced. Three methods of generating holograms (polka dots approximation, DBBI: direct binarisation of interfered beams and optimisation by binary search) loaded onto the SLM to diffract the incident beam are thoroughly described and compared. It is found that in both aberration-free systems and cases with aberrated phase caused by the non-flat SLM, the DBBI method can achieve the highest diffraction efficiency of all three methods. The speed of the optimisation method is the lowest, hence there is no need to extend further. By using the DBBI approach with correction and phase stepping, the diffraction intensity is found to be the most uniform and with the best coverage across the field of view, and therefore the reconstructed SIM images can present a

more compete field of view and fewer artefacts. The role of pinhole size in a SIM system is also discussed. It is found that, although utilising the dots approximation with a doubled pinhole size can also obtain entire FOV coverage, this method cannot obtain as high an accuracy in the phase and amplitude for the carrier terms used by the SIM reconstruction as the DBBI with correction and phase stepping, and so potentially result in the introduction of image artefacts. Accordingly, the DBBI method is found to be the most appropriate approach to generate holograms in SIM systems.

Chapter 6 presents the optimisations made in the current optical systems and platforms. The system can be aligned more easily and conveniently due to the addition of piezo actuators, a piezo stage and optics for phase recovery. The existing ScopeFoundry based programme HexSIMInstrument is optimised to integrate the new hardware and measurements into the system. The implementation of in-situ phase correction and phase stepping are presented. The synchronisation between the camera capturing raw images and the SLM displaying holograms is explained. Due to the phase correction in the DBBI method, the diffracted orders in the HexSIM system are recovered to a diffraction-limited state and pass through the pinholes on the mask completely. Consequently, the raw images with the phase recovery are of an entire FOV, while results without the phase correction have a limited FOV. If the phase stepping is also added in the DBBI method, the sample can be illuminated more uniformly, and there are hence fewer hot pixels in the reconstructed images than the case only with the phase correction. The reconstructed results show the power of the DBBI with phase recovery and phase stepping in enlarging FOV and reducing hot pixels.

The conclusions of all preceding chapters and the future scope of our work can be found in Chapter 7. Our system can be further optimised by introducing beam polarisation controls and integrating more hardware in HexSIMInstrument. A spatial shift will be introduced for each hologram to make the diffracted orders less coherent. The DBBI method with phase correction will be released in a conventional SIM system with sinusoidal structured illumination. In addition, we look forward to further developing the two napari plugins to include more SIM scenarios.

Chapter 2. An introduction to structured illumination microscopy

The goal of this chapter is to briefly introduce the structured illumination microscopy (SIM) technique and to provide the motivation for successive chapters in this thesis. The introduction starts with fluorescence and confocal microscopy as a background. The resolution enhancement and optical sectioning of confocal microscopy are briefly discussed. Then, the concepts of optical sectioning SIM and that of classic SIM (based on the Moiré effect) are introduced. In terms of SIM illumination patterns, conventional sinusoidal illumination patterns and the two-dimensional hexagonal illumination are explained, before illustrating the principles behind the resolution enhancement in SIM systems using these patterns. Finally, the implementation of a real SIM system is outlined, and ideas for optimising the set-up are proposed.

2.1. From epifluorescence microscopy to confocal microscopy

Epifluorescence microscopy is a commonly used technique for imaging fluorescent samples, which is widely used in cell biology, microbiology, and medical diagnostics for its ability to image biomolecules and cells in a non-destructive manner. Confocal microscopy, on the other hand, revolutionises imaging by using a point-like light source and a pinhole aperture to eliminate out-of-focus light. This technique facilitates sharp and high-resolution optical sectioning of samples, which is thus especially suitable for imaging thick samples and realising three-dimensional imaging. Therefore, confocal microscopy is utilised to study structures and dynamic processes at the subcellular level.

2.1.1. Brief introduction of epifluorescence microscopy and confocal microscopy

To observe and study the organisation and movement of molecules within cells, fluorescent labels (e.g. dyes) can be used to label specific cellular constituents. This allows the imaging of specific structures or biomolecules within a sample. Epifluorescence microscopy is an imaging technique which uses a single objective lens to both illuminate (and therefore excite the fluorophores) an entire sample, and also capture the emitted fluorescent light—thereby using a common optical path for both illumination and imaging [1]. As shown in Figure 2-1 (a), a light source passes through an excitation filter and then through a dichroic beam splitter that is

chosen to reflect the excitation wavelength but to allow the emission wavelengths to pass through. This is possible as in general, when a fluorescent molecule absorbs light (i.e. is excited), the emitted fluorescent light will have a lower energy (i.e. longer wavelength) which compared to the excitation light wavelength: a phenomenon known as Stokes shift [2]. The reflected excitation light source is then focused onto a sample by an objective lens. The labelled parts in the sample are excited by the incident light and thus emit fluorescent light, which is collected by the same objective. Then, the emitted light goes through the dichroic beam splitter, an emission filter and is eventually captured by a detector. Epifluorescence microscopy is known as a 'wide-field' imaging technique, the fluorescent molecules from the entire field of view can be imaged (when a camera is used as the detector) as the entire field of view is illuminated.

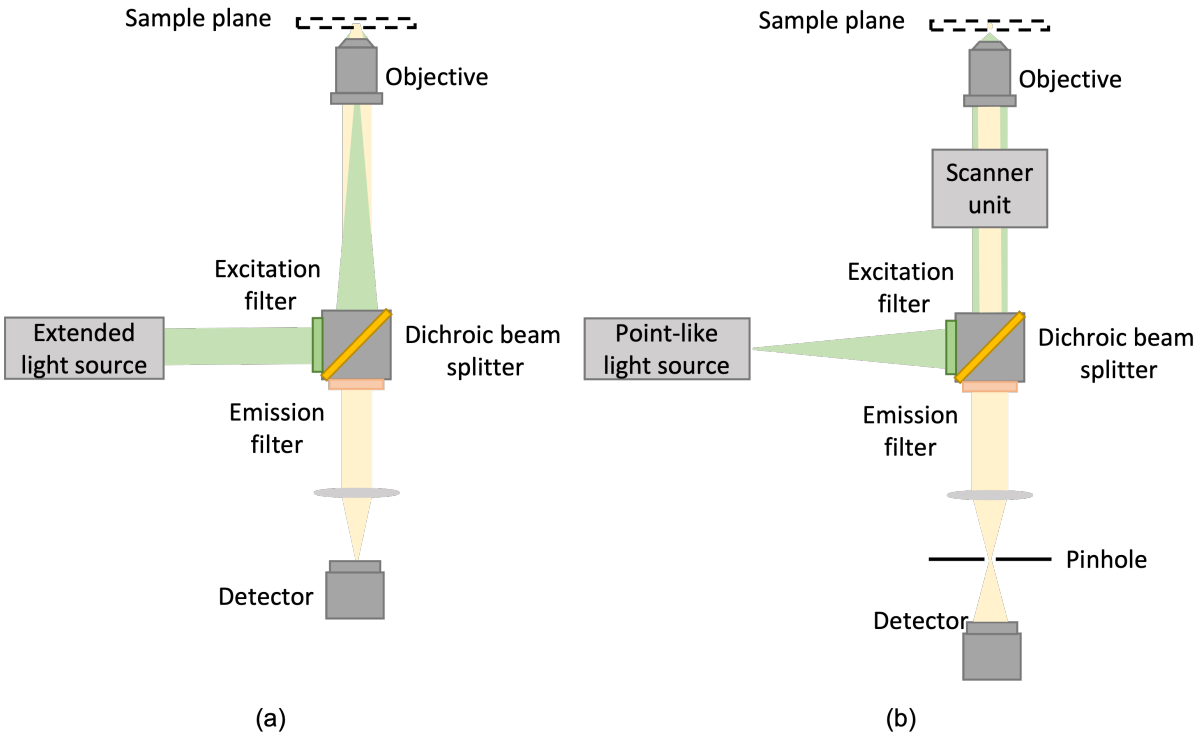


Figure 2-1. (a) Schematic of a typical epifluorescence microscopy system. (b) Schematic of a confocal microscopy system. Relay lenses are emitted in images.

Epifluorescence microscopy is the foundation of many advanced microscopy techniques, however, there is a limit in its capability to detect high frequency details in a specimen. In optical microscopy, the ability to distinguish two adjacent points in a sample is known as resolution. In 1873, Ernst Abbe first introduced that the resolution of a microscope system is intrinsically constrained by a limit [3], now known as the Abbe diffraction limit:

$$d = \frac{\lambda_{em}}{2NA} = \frac{\lambda_{em}}{2n \sin\theta} \quad 2\text{-}1$$

, where d is the shortest resolvable distance and λ_{em} is the wavelength of the emission light. NA is the numerical aperture of the objective lens, depending on the refractive index of the medium n and the half angle θ of diffracted rays. Minsky introduced confocal microscopy, a technique that can be used to improve the resolution of a microscope system [4] [5]. Compared to wide-field microscopy where the whole sample is exposed to the light source, in a confocal system a point-like source, such as a laser or pinhole, is imaged onto the sample and a pinhole is placed before the detector at the image of the point-like illumination to block light emitted from the out-of-focus regions of the sample. Since only a single point is detected at a time, a photodetector may be used to record the signal, a camera is not required. As stated previously, the limitation in optical resolution arises from the mixing of light diffracted from two closely spaced points within a sample. When using confocal microscopy, both light from unfocused planes and from off-axis points can be blocked by the pinhole. The probability of detecting light from a fluorescent point in the sample is a product of how likely it is to be excited by the source and how likely it is to be detected by the pinhole and detector. The two terms are the illumination intensity point-spread-function (PSF) and the detector intensity PSF. With very small illumination and detector pinholes these are both functions of the objective NA and so their product is narrower in size than each individually. The higher diffraction features of a point seen in a conventional microscope, both out-of-focus and within the focal plane, can thus be eliminated (as their intensity is squared), and the contrast in the image is thus higher. Because light that is not from the focal plane is largely removed, a confocal system can perform optical sectioning. By itself, a confocal arrangement detects a single image point, but if a scanner unit is added in a confocal system to move either the sample or the incident beam, the fluorescent molecules can be detected as a 2D image. Then 3-D imaging can be realised by scanning the sample volume through the image plane (or vice-versa) and reconstructing the volumetric data. A diagram of a confocal microscope is presented in Figure 2-1 (b), where both the illumination and imaging paths, with a pinhole located at the focal plane before the detector, are imaged by the same objective lens onto the same point in the sample and these conjugate light paths are controlled by a scanner unit to scan the image point across the sample in order to generate a 2D image. As a point source is required, this often achieved using a laser, which limits the range of available wavelengths in addition to adding significant complexity and cost to early systems.

Although a high resolution and a high contrast can be achieved in a confocal microscopy set-up, the price is a low signal-to-noise ratio (SNR) since the emission fluorescence light is largely

blocked [6] if a small pinhole is used. Furthermore, as the system has only a single point detector, image acquisition times can be very long for 2D and 3D samples. It may be noted that there are schemes to parallelise acquisition in confocal microscopes in order to reduce acquisition times, such as slit and spinning disk microscopes, but these come at the cost of increased cross-talk between pinholes that reduces image contrast [7] [8] [9].

2.1.2. Resolution in conventional and confocal microscopy

Resolution in an objective lens is limited by Abbe diffraction limit, which depends on the lens' numerical aperture. Namely, the detectable frequency of a microscope objective is restricted in a pupil with a radius as the cut-off frequency in the reciprocal domain. In practice, the lens produces a distribution of intensity in its focus even when a perfect point is being imaged [10]. To analyse this, the Fourier transform is used to associate the resolution in the real space and the observable frequency in the Fourier (frequency) space. In the real domain, the point spread function (PSF) that describes the response of a point source is used to define the resolution in a microscope system. The complex field PSF is the inverse Fourier transform of the pupil function, and the Fourier transform of the intensity PSF is the intensity optical transfer function (OTF). As shown in Figure 2-2 (a), the intensity OTF support of a conventional microscope is of a torus shape [11] [12]. The missing cone [13] present in the OTF at low spatial frequencies hinders the wide-field microscope from realising true three-dimensional imaging as it cannot measure any information from the sample as to how these low spatial frequencies change with defocus.

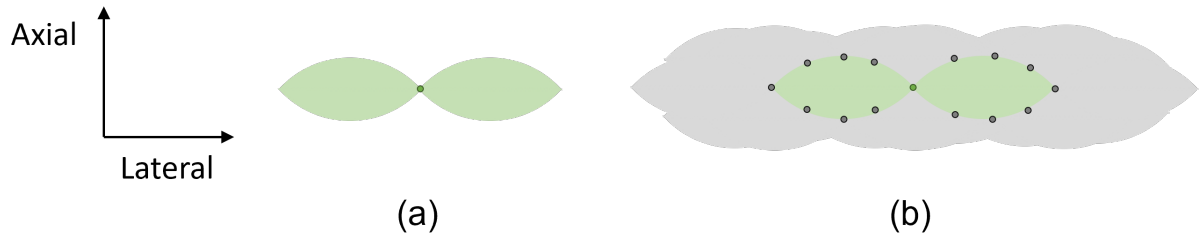


Figure 2-2. The OTF support in (a) conventional microscopy and (b) confocal microscopy, respectively. (b) shows the extended OTF support generated by the convolution of two supports in (a). The centre of each OTF support is denoted using a circle. In the schematic, infinite numbers of OTF supports are simply shown as 14 supports (see grey centres).

While in a confocal microscope system, the effective PSF is the product of the illumination and the detection PSFs. For a small detection pinhole the two PSFs are approximately the same,

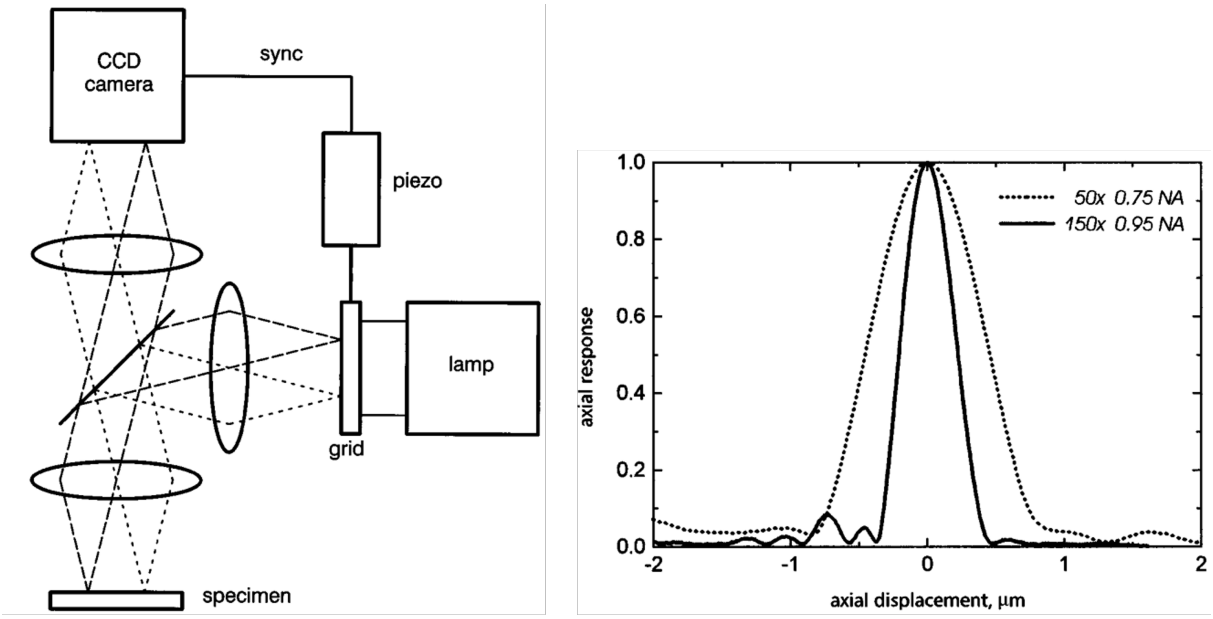
there thus is a roughly $\frac{1}{\sqrt{2}}$ decrease in FWHM of the PSF. Correspondingly, as presented in Figure 2-2 (b), the OTF in confocal microscopy can be seen as the convolution of two conventional OTFs. The region of OTF support is doubled and the missing cone issue is addressed. Compared to conventional microscopy, the resolution is enhanced, and the out-of-focus blur is removed.

2.2. Structured illumination microscopy

Confocal systems are different to wide-field microscopes due to the introduction of the pinhole that results in optical sectioning and resolution beyond the diffraction limit. However, confocal microscopy is not the only technique to get around the diffraction limit, there are lots of techniques have been developed, each utilising a different mechanism to overcome the limit. Some include: exploiting a nonlinear dependence of the fluorescence emission rate on the illumination intensity (nonlinear SIM) [14]; exploiting localisation of single fluorophores, as an example, stochastic optical reconstruction microscopy (STORM) [15] [16] [17]; modulating the PSF such as stimulated emission depletion (STED) [18] [19] [20]. Among these, structured illumination microscopy (SIM) microscopy is a reasonable method to surpass the diffraction limit without compromising the SNR and the imaging time, although the extent of resolution improvement is limited. Two essential features of SIM microscopy: the low phototoxicity and the doubled resolution in three dimensions make it favourable for imaging live cells and their activities.

2.2.1. Concept of SIM for optical sectioning (OS SIM)

In 1997, Neil *et. al* applied structured illumination in a conventional microscopy system to produce optically sectioned images [21]. Based on conventional microscope, as presented in Figure 2-3 (a), the microscope's illumination system is modified to project a grid pattern onto the object.



(a) (b)

Figure 2-3. Schematic of the optical layout. (b) Axial response measurements in the system. Reprinted with permission from [21] © Optical Society of America.

As images are acquired in the microscope the grid pattern is only visible on the in-focus parts of the specimen, which can then be identified as the grid is moved. The intensity of the image is calculated with respect to the grid pattern. By taking three images, I_1, I_2 and I_3, with the grid pattern at different relative spatial phases (0, $\frac{2\pi}{3}$, and $\frac{4\pi}{3}$), the grid pattern can be removed, by a simple calculation of the mean square of the three differences between the images, resulting in an optically sectioned image:

$$I_{sect} = \sqrt{(I_1 - I_2)^2 + (I_2 - I_3)^2 + (I_3 - I_1)^2} \qquad 2\text{-}2$$

. Meanwhile a conventional non-sectioned image can be calculated from:

$$I_{conv} = I_1 + I_2 + I_3 \qquad 2\text{-}3$$

. These simple calculations enabled video-rate images to be produced that had similar optical sectioning properties to those of confocal microscopes without the added complexity. The measured axial responses when imaging a mirror surface for lenses of two different NAs are shown in (b). This type of method is referred to optical sectioning SIM (OS-SIM), and has been further developed for fluorescence applications [22]. M. Neil and T. Wilson obtained real time optically sectioned fluorescence images in a widefield conventional microscope by the interference of two beams on a specimen [23]. Karadaglić proofed that by using structured illumination, optical sectioning in conventional bright-field reflection microscopes can be

comparable to results of confocal microscopes [24]. Karadaglić and T. Wilson then analysed structured illumination wide-field fluorescence microscopy (SIWFFM) and demonstrated that both optical sectioning and the transfer function behaviour of this technique can be as same as confocal microscopy [22].

However, the method has a number of drawbacks. Firstly, the grid is projected onto the sample by an incoherent imaging process. In some respects this has the benefit that a laser is not required meaning that the system is cheaper and more flexible than laser confocal systems. It also benefits from the fact the illumination is optically sectioned as well as the imaging path, as the contrast of the projected grid pattern decays rapidly with defocus because of the incoherent imaging. The major drawback though of using incoherent illumination is that the contrast of the projected grid pattern also attenuates quickly as the spatial frequency of the grid is increased. For this reason, OS-SIM systems regularly use quite coarse grid patterns, which as discussed below are not helpful in increasing resolution. Finally, the calculation in equation 2-2 is inherently non-linear and so leads to artefacts such as the rectification of noise from out-of-focus light. This out-of-focus light is also something that would have been physically blocked by the pinhole in a confocal microscope and therefore would not have contributed any noise to the final image.

2.2.2. Enhancing lateral resolution with SIM

In 1994, Bailey *et. al* generated a standing wave illumination by interfering the light passing through two high-NA objective lenses, enhancing the axial resolution [25]. The concept of using illumination patterns with a sinusoidal intensity structure to improve the lateral resolution in fluorescence microscopy was introduced later by Heintzmann in 1999 [26]. After one year, Gustafsson *et. al* illuminated fluorescent microspheres with sinusoidal patterns to obtain a two-fold lateral resolution enhancement and discussed SIM's extension to three dimensions [27]. This is regarded as the first paper to explain principles behind SIM microscopy and to demonstrate the resolution improvement in reconstructed images. In 2008, the validity of 3-D resolution doubling was confirmed theoretically and experimentally [28], where the method for reconstructing high resolution images was also explained in detail.

In SIM microscopy, instead of a uniform illumination, a structured illumination pattern is projected through the objective into the specimen to extend the resolution of a wide-field microscope system, as will be explored in the following sections. The most common pattern is

generated by the interference between two coherent beams, proceeding from the objective lens, with the pattern illustrated locating at the object plane.

Having covered the practical realisation of structured illumination pattern generation, it is worth introducing the reasons why illuminating a sample using structure light patterns can overcome the diffraction limit. The fundamental principle behind the high resolution of SIM is the Moiré effect [29], which illustrates that a low frequency can be generated by superposing two high frequencies. An example of Moiré patterns observed in daily life is shown in Figure 2-4 and is indicated using arrows. When a fine structure with a high frequency (see a) at the front (or back) of a chair's back is overlapped with another high-frequency pattern at the back, the mixed fringes demonstrated in (b) are much coarser than the pattern in (a). These are called Moiré fringes. Similarly, in a SIM microscope system, a sample with unknown frequencies which are higher than the maximum detectable frequency of the wide-field set-up is illuminated with a known high-frequency pattern. As a result, the high frequencies of the sample can superimpose with those of the (known) structured pattern, and the Moiré fringes from these overlapped patterns are low enough to be detected. The illumination patterns can be rotated to provide low-frequency information from high frequency components in different directions in the sample. The illumination patterns can be shifted so that the information about the fine structured in the sample encoded at low frequencies into the raw images captured by the SIM system can be decoded through reconstruction.

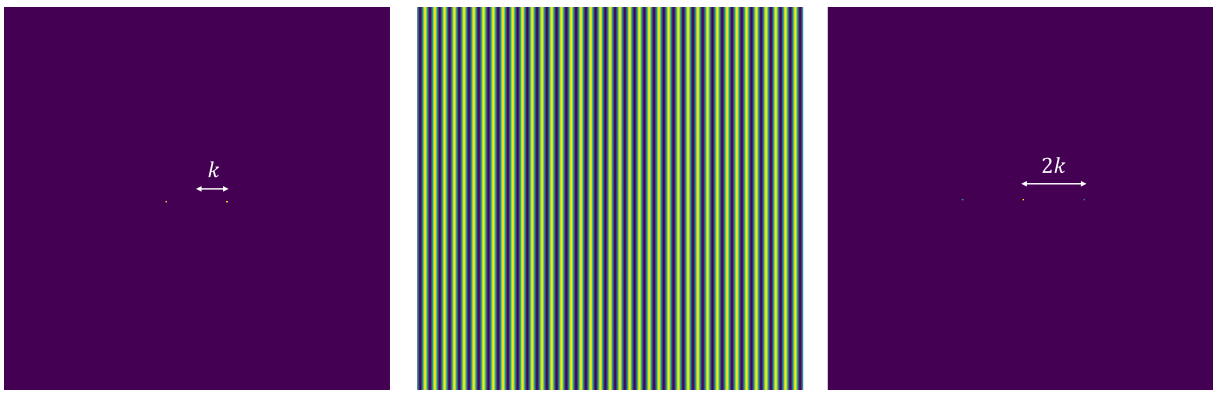


Figure 2-4. The Moiré effect in daily life. (a) demonstrates the pattern of small round wholes at the front and the back of a chair. (b) The overlapped patterns of both sides of the chair's back are Moiré stripes, which are denoted by arrows. Coarse stripes can also be observed from yellow chairs.

2.2.3. SIM with sinusoidal illumination

Before explaining specifically how SIM microscopy surpasses the resolution limit in conventional microscopy, the two illumination patterns previously outlined will be characterised to facilitate the determination of the maximum resolution extension factor, which will be used in the following section.

SIM illumination patterns are commonly generated by the interference between multiple beams. The most common SIM set-up is to interfere two beams to produce a sinusoidal illumination pattern. This can be interpreted using Fourier theory. As illustrated in Figure 2-5 (a), a field of two beams (a) is simulated, with the distance between a beam and the centre of the field being a wavevector shown as k_x, but in reality this corresponds to the beams being focussed at positions near the edge of the pupil of the objective lens. The field undergoes a Fourier transform and then multiplies with its conjugate to obtain the illumination intensity as shown in (b). This is the result of the interference between the two beams. To discover the relationship between the frequency components of the interference of the two beams, the patterned illumination intensity is subjected to an inverse Fourier transform, as shown in (c). There are three frequency components generated by the interfered beams, including two that are conjugate to each other and a DC frequency. Geometrically speaking, the frequency components of the illumination pattern are twice those of the beams and so if the beams were initially placed at the edge of the pupil they will have $k_x = k_0 NA$, where k_0 is the wave number ($\frac{2\pi}{\lambda}$). As shown in (c), the distance between a frequency of the illumination and the DC component is $2k_x = 2k_0 NA$, which is the maximum spatial frequency that the lens can transmit in an incoherent image. Using this sinusoidal pattern enables frequencies up to $4k_x = 4k_0 NA$ to be seen by the microscope because of the Moiré effect and so gives a rise of double the detectable frequency compared to that in the conventional case.



(a) (b) (c)

Figure 2-5. (a) The field amplitude of two beams. (b) The intensity of the Fourier transform of the two beams. (c) The interference between two beams. Distances between components and the centre are denoted using arrows. All complex fields are shown only with real amplitude values in the figure.

For conventional microscopy, in reciprocal (spatial-frequency) space, the highest transmittable frequency is defined by the diffraction limit as the cut-off frequency $k_c = 2k_0NA$,. As shown in two lateral dimensions in Figure 2-6 (b), the observable region in the frequency space is determined as a circular area with the radius as k_c. The area can also be regarded as the support region of the optical transfer function (OTF) of the system. When a sample with a spacing as $\frac{2\pi}{k_{sample}}$ is illuminated by a sinusoidal light pattern with a frequency as k_{ill}, as demonstrated in (a), Moiré stripes with a spacing that is coarser than $\frac{2\pi}{k_{sample}}$ are generated by the frequency mixing. The mixed frequency is denoted as k_M, fusing frequency information of k_{ill} and k_{sample} together. In fact, the positions of the two frequency components of the structured illumination in the reciprocal space is determined by the spacing of the pattern.

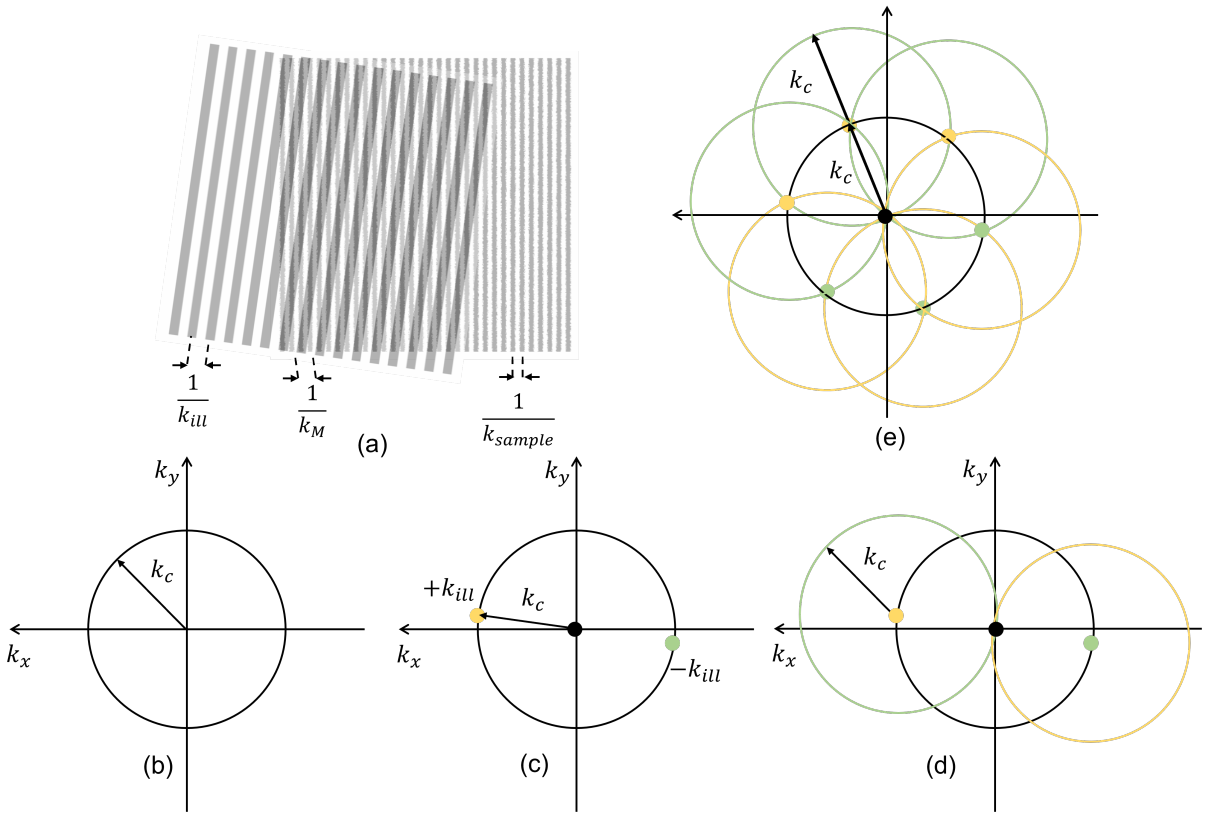


Figure 2-6. Principle behind resolution enhancement in conventional SIM microscopy. (a) The sample is superposed with the structured illumination pattern, where frequencies are denoted. (b) The detectable area of the conventional microscope system in reciprocal space, which is can also be seen as the OTF support region. (c) Two frequency components of the illumination pattern are added into the observable area and denoted in green

and yellow. The zero-frequency component located at the centre is shown as a black spot. (d) The two frequencies extend two offset regions based in reciprocal space. (e) The detectable region is expanded to a twice size by rotating the structured illumination pattern.

As shown in (c), the yellow and the green spots represent the two illumination frequencies, and the DC component that locates at the centre, denoted in black, when the sample is illuminated as in (a). The spatial-frequency of the illumination (k_{ill}) is usually set as high as possible to maximise the observable region. The maximum k_{ill}, however, is also constrained by the diffraction limit in epi-illumination, as the light path of the emission and the illumination share the same objective, limiting it to a maximum equivalent to k_c. Actually, the superposition of two patterns shown in (a) is equal to the sample signal multiplied by the illumination one. In Fourier space this is equivalent to the convolution of the Fourier transform of the two signals. Therefore, as illustrated in (d), compared to the observable region of the system, the convolution with the yellow spatial frequency component shifts the spatial frequencies in the yellow circle with the same radius k_c into the black circle where they can now be seen by the objective, and similarly for the green spatial frequency component and the green circle. Meanwhile the black DC component results in a conventional image. The maximum frequency value in the support region is the sum of $k_c$and k_{ill}, being theoretically $2k_c$. Accordingly, the detectable frequency is enlarged by a factor of two, extending the observable region in one direction. To enlarge the detectable spectra in various directions, the structured illumination pattern is step-rotated [30], as illustrated in (e), resulting in an observable area that is twice the size than that in the wide-field case. By applying illumination patterns, the support region is expanded twofold to detect high-frequency information, thus doubling the potential spatial resolution compared to a conventional system.

In three dimensions the enhanced region of support can be visualised as the combination of shifted conventional OTFs (Figure 2-7 (a)) to the carrier positions as shown in Figure 2-7 (b). The separated components are multiplied with the OTF [31]. Note here that the penalty of using illumination carriers close to the incoherent cutoff is that the resulting 2-beam 3D-OTF still has missing cones now additionally at the illumination frequencies as well as at the DC frequency. This can result in artefacts that make it often preferable to use illumination spatial frequencies below that of the cutoff, $k_{ill} < k_c$.

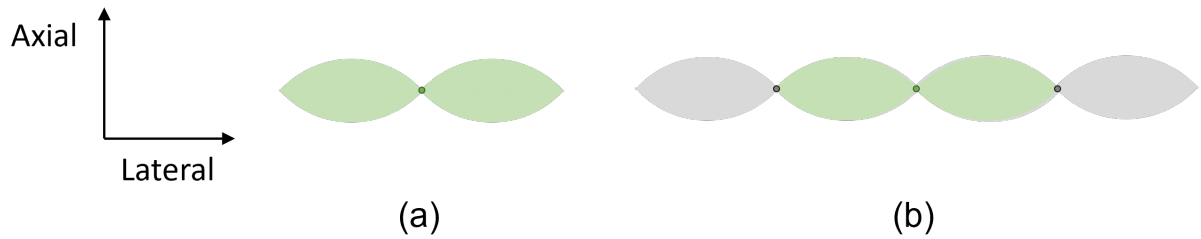


Figure 2-7. The 3D OTF in (a) conventional microscopy and (b) a two-beam SIM microscope. The shifted OTFs are in grey. The centre of the OTF is denoted using a circle.

However, the overlapped circular regions in the frequency domain are fused and consequently need to be separated and shifted back to their correct spectral positions. To separate the three frequency components in Figure 2-6 (c), three images of the specimen illuminated by patterns with three phases are collected. At each phase, the illumination pattern is required to be rotated to three orientations to obtain a more isotropic spectral extension, resulting in three in-phase images with three orientations [32]. Therefore, nine raw images are collected and then reconstructed into an enhanced resolution image. With the decoded higher frequencies, the resolution in the reconstructed images will be double the conventional wide-field resolution.

2.2.4. SIM with hexagonal patterned illumination

Similar to the case with 2-beam SIM, the relationship between the spatial frequencies of a hexagonal illumination pattern and the locations of three beams can be explored. As demonstrated in Figure 2-8 (a), three beams are simulated with the spacing between a beam and the centre of the field as k. The Fourier transform is applied to the three-beam field, and the result is multiplied with its conjugate to generate the hexagonal structured pattern in (b). Then, the inverse Fourier transform of the illumination pattern, as illustrated in (c), reveals the magnitude of the frequency component of the intensity of the interfered beams is $\sqrt{3}$ times that of the original field of the individual beams, corresponding the the difference in the beam spatial frequencies. The locations of the six carrier frequency components of a hexagonal illumination will be at $\frac{\sqrt{3}}{2}k_c$ from the centre in the Fourier space of the structured pattern.

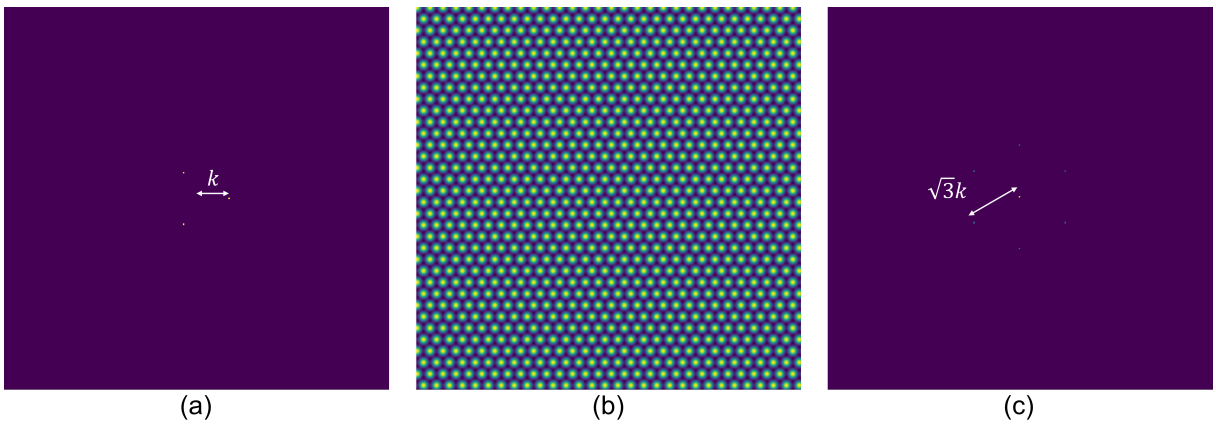


Figure 2-8. (a) The field of three beams. (b) The intensity of the Fourier transform of the three beams. (c) The inverse Fourier transform of the interference intensity. Distances between components and the centre are denoted using arrows. All complex fields are shown as amplitude values in the figure.

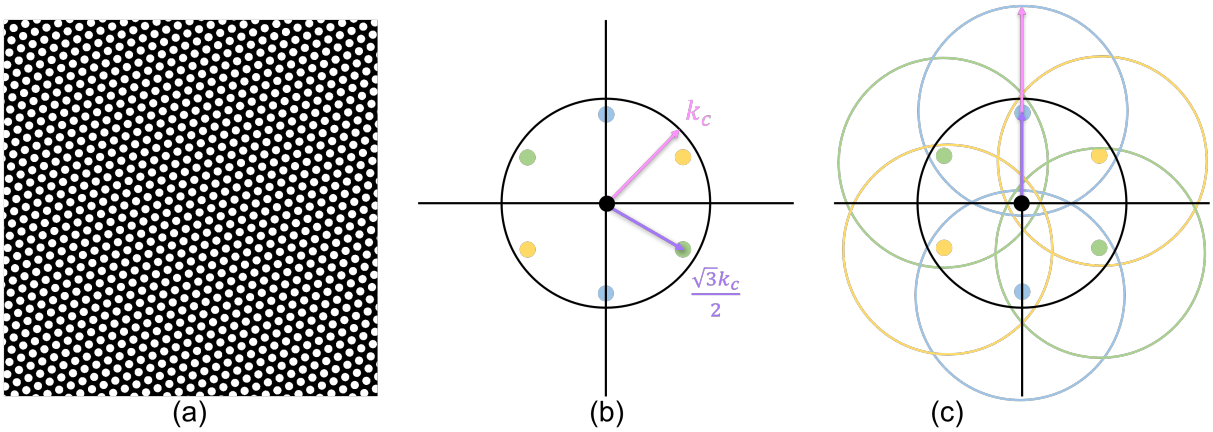


Figure 2-9. The resolution improvement in HexSIM microscopy. (a) A hexagonal mask for generating a hexagonal illumination pattern. (b) The detectable area of the conventional microscope system in reciprocal space and frequency components of the hexagonal illumination. The zero-frequency is denoted as a black spot, while the other three are shown in blue, green and yellow. (c) The detectable region is enlarged by applying the structured illumination pattern without rotations. Conjugates of frequency components and the DC component are omitted in the figure.

In addition to the conventional stripes generated by two beams, hexagonal patterns generated by three beams can also be used to outperform the resolution limit. As can be seen in Figure 2-9 (a), a hexagonal illumination mask is used consisting of polka dots. This mask generates 6 first order diffracted beams in a hexagonal arrangement, but if only three of these beams are selected, by masking in the pupil plane, the resulting intensity pattern consists of three pairs of frequency components in reciprocal space (b) in addition to the zero-frequency. Obviously, we don't need to rotate the hexagonal illumination pattern because the region of support can be extended along the orientations of these three frequency components. Thus, only seven raw images with different phases need to be collected to reconstruct a super-resolved frame when hexagonal illumination [33] is applied. As explained in 2.2.3, the distance between the carrier frequency components and the centre is a maximum of $\frac{\sqrt{3}}{2}k_c$ when the incoherent cut-off frequency of the microscope system is k_c. As illustrated in (c), the detectable region can be correspondingly expanded without rotating the hexagonal pattern because of the three frequencies. Correspondingly, the factor of resolution enhancement is $1+\frac{\sqrt{3}}{2}\approx 1.87$. In our system, we adopt hexagonal illumination patterns such that fewer raw images are required to reconstruct a super-resolved image and to avoid rotating illumination. The advantages and the design of hexagonal structured illumination are discussed further in Chapter 5.

Heintzman *et. al* compared many illumination patterns that can be used for SIM in terms of signal-to-noise ratio and resolution in both linear and nonlinear SIM systems [34].

2.2.5. SIM reconstruction algorithms

When the sample in a SIM system is illuminated with structured patterns, images are collected (known as raw frames) that include encoded high frequency information, undetectable by a wide-field system. The raw images are reconstructed to extract the high-frequency information and reconstruct a super-resolved image. Reconstruction algorithms are essential in this technique, and their development has thrived over the past decade.

In 1999, Heintzmann *et al.* proposed that raw images can be processed based on the spectrum of a sample in the Fourier space [26], and the technique was called as laterally modulated excitation microscopy (LMEM) at that time. Gustafsson *et al.* introduced the generalised Wiener filtering SIM reconstruction method in 1999, with a conventional SIM system using a grating to generate sinusoidal patterns [27]. In 2008, the method was extended to surpass the diffraction limit in both lateral and axial directions to realise 3-D SIM, using an algorithm unveiled by Gustafsson *et al.* [28]. Lal et. al [35] presented further theoretical details and an open source SIM reconstruction algorithm (OpenSIM). A practical protocol for SIM has been introduced by Demmerle *et. al* [36], including the acquisition of calibration parameters, artefact correction and even preparation of calibration bead slides.

In addition, Schropp *et. al* have developed reconstruction algorithms for systems using hexagonal patterns and other 2-D patterned illumination in [33] and [37]. The reconstruction procedure in this thesis is based on the algorithm introduced in [38], where seven raw images are processed by initially pre-filtering and up-sampling the raw images, and then unmixing and remixing the frequency carrier components within the images in the spatial domain, before a Wiener-filter is applied to generate the resulting reconstructed image. There are of course many more reconstruction algorithms than those covered in this brief overview, a more detailed investigation into reconstruction algorithms, beyond the scope of this thesis, can be found in [39].

2.3. Optimisation of an existing SLM HexSIM system

The implementation of our SIM system using hexagonal illumination and a ferroelectric liquid crystal spatial light modulator (FLCSLM-HexSIM) is based on the set-up introduced in [38], as

shown in Figure 2-10. To optimise the performance of the system, a napari plugin was developed to simulate this system, and the plugin was extended to cover various SIM configurations, which are presented in Chapter 3. Practically, there was an issue with the installation of the SLM: the SLM's surface is not flat, which leads to aberrations in the phase of the SLM and resultant extended diffraction orders being located at the spatial filter shown in the figure below. This was addressed by applying a phase-recovery algorithm with stepped phases, which will be introduced in Chapter 4. The method of designing hexagonal patterns is optimised and illustrated in Chapter 5, making the phase correction accessible without the compromission of the uniformity of illumination patterns. The optimisation in terms of hardware will be introduced in Chapter 6. The spatial light modulator (Forth Dimension Display, SXGA-R3) with a graphical card driver is replaced with a new model (Forth Dimension Display, 2K-R11). The liquid crystal screen of the new SLM has more pixels and a smaller pixel size, providing a larger field of structured illumination and more accurate control of artefacts. Besides, more triggering options are available using the new SLM with a USB driver. To improve the safety and facilitate the remote alignment of the system, both the new SLM and the dichroic beam splitter are connected to a piezo stage.

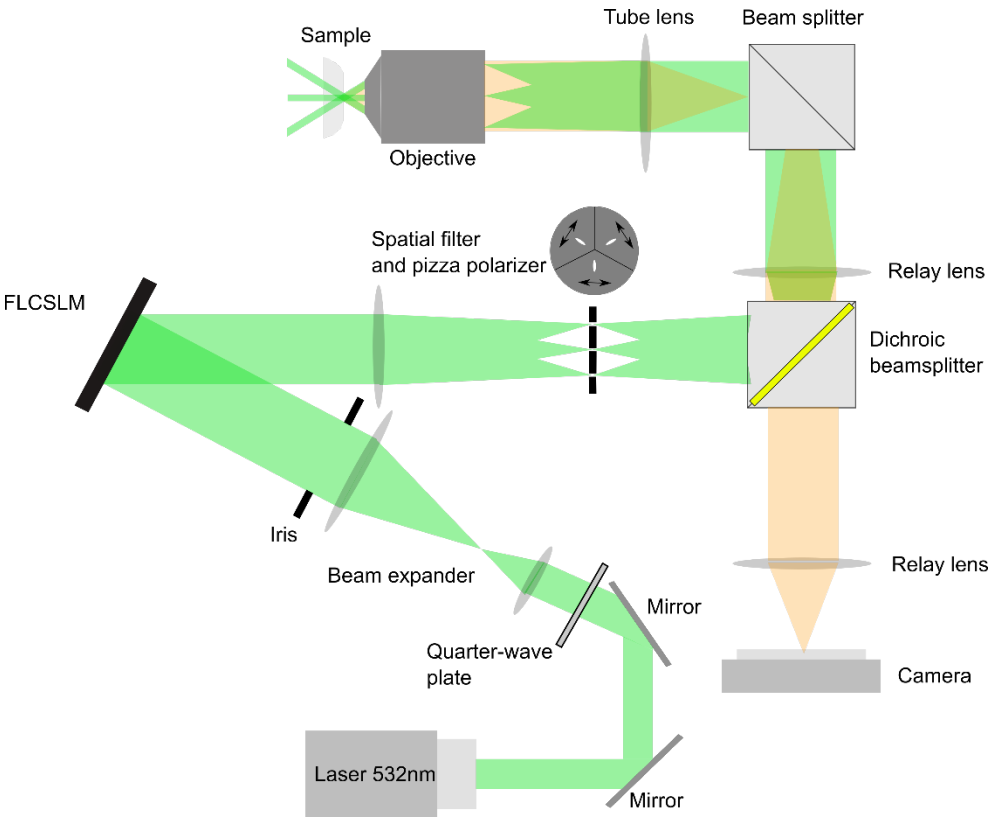


Figure 2-10. The set-up of the previous HexSIM system based on FLCSLM. Figure is adopted from the supplementary document of [38] with permission.

2.4. Increasing resolution further in optical microscopy

While SIM is limited in the maximum spatial frequency that can be achieved, other techniques such as (nonlinear SIM, STORM and STED) are not. Similar to typical SIM, nonlinear SIM ([40] [41] [42] [43] [44]) also utilises structured illumination. However, this technique can theoretically push the resolution limit to infinity, although the signal-to-noise ratio restricts this in practice. Nonlinear responses of fluorophores to the patterned illumination are exploited to create higher harmonic terms in the effective illumination to increase the resolution further, though requiring more raw images to do so.

Single molecule localisation microscopy (SMLM) methods such as stochastic optical reconstruction microscopy (STORM) ([45] [16] [46] [17] [47] [48]) on the other hand surpasses the Abbe resolution limit, by making the assumption that a single emitting fluorophore can be localised on a smaller scale than the optical PSF of an imaging system, based on the prior knowledge that the detected PSF image comes from just that single emitter. The sample is prepared in such a way that a laser causes fluorophores to switch between dark and fluorescent states, such that at any given time, only a subset of fluorophores is activated. Based on the image of these sparse fluorophores, the locations can be accurately determined as being at the centre of each measured PSF. Practically, this is done by fitting a function (e.g. an Airy disk or Gaussian model) to each PSF to ‘localise’ the fluorophore position. Given ideal conditions, the resolution can be reduced down to 20-50 nm [49]. The precision of the localisation of each fluorophore is dependent on both the size of the optical PSF and the number of photons recorded however [50]. The switching of random different sets of fluorophores and their subsequent localization needs to be repeated many times to create a super-resolved image, a STORM acquisition therefore typically requires thousands of frames to be recorded, leading to long acquisition times. Furthermore, the use of ‘switchable’ fluorescent dyes, which can be activated and deactivated requires particular sample preparation, including the use of chemical buffers, and therefore not all fluorescent dyes are suitable for use in STORM imaging.

Stefan Hell received the Nobel Prize (along with Eric Betzig and William Moerner) in Chemistry in 2014 [51] for developing STED microscopy, which has significantly advanced the field of optical microscopy. In STED, two lasers, an excitation laser and a depletion laser, are exploited. The intensity profile of the depletion laser is of a donut-like shape, and it depletes the fluorophores in the periphery by simulating the excited molecules to return their ground

state. Therefore, only a small central region of fluorophores remains not depleted. By shrinking the effective size of the fluorescent spot via stimulated emission depletion, STED microscopy achieves spatial resolutions well beyond the diffraction limit, typically reaching 20-30 nanometres. STED ([19] [52] [53]) microscopy provides a substantial enhancement in resolution compared to conventional microscopy, enabling the visualisation of structures at the nanoscale. It has also been used to perform super resolution imaging in the axial direction [54]. Furthermore, it can be combined with multi-photon and confocal imaging and is therefore suitable for imaging relatively deep into tissues, it has even been demonstrated on *in vivo* samples [55]. However, as it is a laser point scanning technique, it features long acquisition times.

2.5. Chapter summary

To summarise, this chapter provides the background of SIM microscopy, providing a foundation for the following chapters of this thesis. Wide-field microscopy was introduced and was compared with confocal microscopy, which uses a pinhole to restrict the field of imaging. SIM microscopy is a wide-field (epifluorescence) technique, which is originally used to enhance optical sectioning and called optical sectioning SIM. It has been developed to surpass the diffraction limit by a factor of two. The technique is based on the phenomenon where coarse stripes can be generated by superposing two fine (high spatial frequency) striped patterns, which is called the Moiré effect. We introduced the two structured illumination patterns. The most common pattern is sinusoidal stripes, which can double the diffraction resolution limit by collecting and reconstructing nine raw images. While the application of the hexagonal illumination pattern (the other common pattern) only requires seven raw frames for per reconstructed frame, the factor of the resolution enhancement is only about 1.87. The resolution limit is transcended by encoding and decoding high frequencies which are not detectable in a wide-field system, namely, extending the detectable support region in the reciprocal space. After introducing the relevant theory, practical optimisations of the previous FLCSLM HexSIM system are announced, including the simulation of SIM systems, the phase -correction algorithm with the phase stepping, the improvements in hologram design and changes in implementations of the HexSIM system. Finally, some super-resolution techniques are briefly reviewed, including nonlinear SIM, STORM and STED microscopy.

Chapter 3. Theory and simulation of SIM imaging

At its most simple, SIM imaging is performed by applying a set of structured light patterns to illuminate the sample and then collecting a matched set of raw images. Therefore, to analyse the performance of a SIM system, the entire imaging process can be simulated by calculating the structured illumination, applying it to phantom samples, and then modelling the imaging process for light emitted by the samples. To do so, we will start with foundational theory of the field near the focal plane of a lens, and then apply it into the imaging of fluorescent molecules in the simulation of a SIM system. The modelling of the imaging system point spread function (PSF) and the optical transfer function (OTF) are first modelled in the scalar case, and then two vectorial models, with fluorescence molecules modelled as both flexible and rigid dipoles, are used. Samples used in the virtual SIM system are simulated as point clouds of spheroids and filaments. Finally, the structured illumination patterns are calculated based on the theory of Jones vectors and then applied to the simulated samples. In this way, complete 'raw' SIM images are simulated. The whole procedure (optical imaging simulation, test sample generation and application of structured illumination) is implemented and packaged as a Python-based plugin for the napari multi-dimensional image viewer system [56]: napari-generic-SIMulator [57]. To analyse the generated structured illumination raw images and to obtain their reconstructions, stacks of raw images are processed by a separate napari plugin: napari-SIM-processor [58]. This chapter presents the theory of optical modelling, as well as discussion of its practical implementation in the napari-generic-SIMulator, demonstrating results of the simulation in each relevant section.

3.1. Simulating image formation in a SIM microscope

Figure 3-1 shows a simplified SIM microscope optical setup, featuring an illumination path containing an illumination lens that produces a number of focussed diffracted orders in the pupil of an objective lens. In the systems described in this thesis, the diffracted orders are produced by displaying holographic gratings on a spatial light modulator (SLM) that is illuminated by a collimated coherent light source. The objective lens re-collimates the diffracted orders so that they interfere in the sample plane to produce a structured illumination pattern. A fluorescent sample will absorb light and re-emit it at a longer wavelength. The objective lens then collects that light and presents it as a field in its pupil plane. Finally, the tube lens produces a focused image of the collected light at the camera plane.

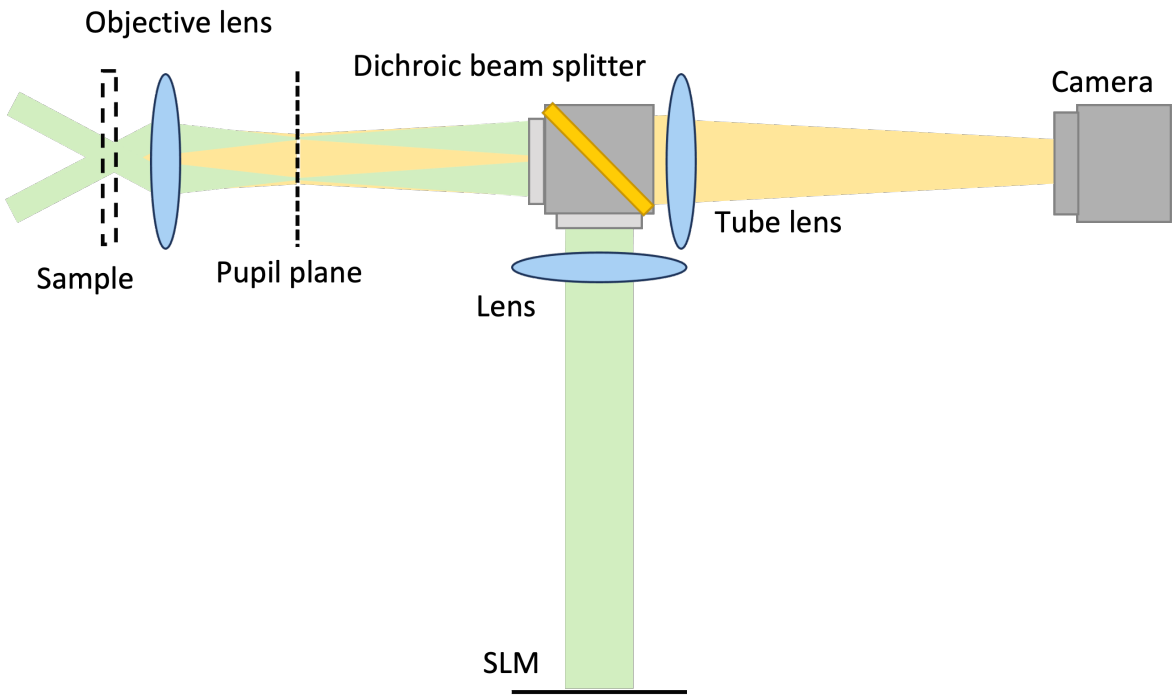


Figure 3-1. Schematic of a SIM system. The light path of the illumination and the emission is demonstrated in green and yellow, respectively.

Each of these processes can be understood in terms of the optical principles of interference and diffraction. However, these familiar processes are further complicated by polarisation effects, particularly on the sample side of the objective lens where waves are travelling at a significant angle to the optic axis. The diffracted illumination orders in the objective pupil plane are produced by diffraction from the SLM in combination with the illumination lens to bring them to a focus, and this is dealt with in much greater depth in Chapter 5 and Chapter 6. The interference of the re-collimated diffraction orders in the sample plane is described here in section 3.2 and includes the analysis of how the illumination pattern polarisation and contrast varies across the sample plane. Polarisation also plays a significant role in absorption and emission in the fluorescence process and is covered in section 3.3. When the objective collects light from the sample a further polarisation rotation occurs as the rays from high angles are brought together in the pupil plane, as covered in section 3.4. Finally, an image is produced in the camera plane that is dependent on both the limiting size of the objective pupil, and the polarisation of the fields collected from the sample found there. We consider how this image system images a point object in the sample plane and so calculate a point spread function or PSF for the system.

The fluorescence process is an incoherent one that can be treated as though we are imaging each individual fluorophore as a separate point source. In fact, we model this as though the

sample consists of a collection of point sources - a point cloud - each of which experiences an illumination field at its location driving its emission. The image of that emission is then calculated as an incoherent convolution with the point spread function of the imaging system taking account of any polarisation effects. The simulation is packaged as a plugin for the napari system that can simulate different illumination patterns and polarisations, different models for the way polarisation is affected by the fluorescence process and different sample point cloud scenarios. Motion of the sample point cloud is also included to simulate real world issues such as Brownian motion and sample drift.

3.2. Calculating the illumination intensity in the sample space

The illumination patterns in SIM can be seen as an electrical field generated by the interference between multiple fully polarised coherent beams. However, for high numerical aperture lenses, it is important to consider how polarised light changes as it is focussed by the lens. O'Holleran and Shaw discussed impacts of the polarisation state on the contract of a structured illumination pattern and presented that both the NA of the objective lens and the electric field's orientation can affect light throughput and the speed of imaging [59]. Figure 3-2 shows the illumination beams in various SIM illumination geometries in terms of how they illuminate points in the objective pupil and how they are recollimated to beams that interfere in the sample plane. The fact that these beams are bent to travel at a significant angle to the optic axis means that we have to be careful of both the polarisation of the field in each beam and how it is rotated by the lens. Jones vectors can be used to represent an optical field, as they describe both the amplitude and the phase so can be employed where calculation of interference between beams is required. However here we are interested in the interference between beams that are travelling at significant angles to each other as well as to the optic axis. Therefore, we use three-dimensional Jones vectors to describe the optical field as they can describe polarisation components in all three directions.

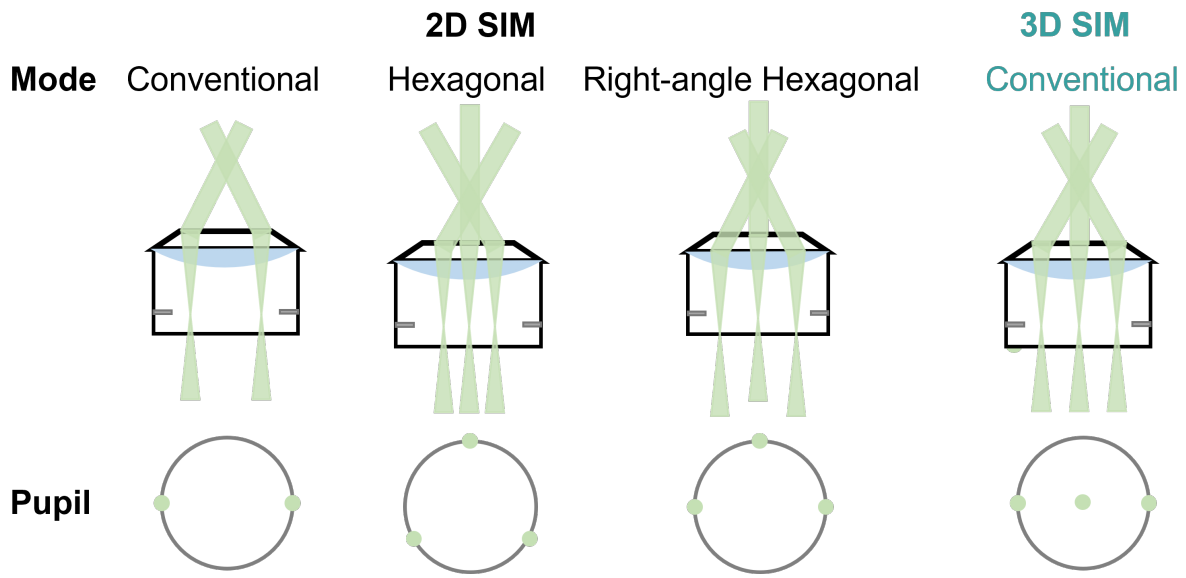


Figure 3-2.Beams in different SIM modes. The grey circle denotes the illumination pupil.

3.2.1. Polarisation rotation by the objective lens

When a light ray changes direction as it is focussed through a lens, the change in its polarisation can be calculated according to two guiding principles. First, the azimuthal component of the optical field that is tangential to the pupil edge is unchanged. Second, the radial component in the pupil is rotated to that still radial with respect to the pupil, but orthogonal to the direction of propagation of the beam after the lens. Therefore, to propagate a general polarisation from a ray in the input pupil to the output space of the lens it should be decomposed into radial and azimuthal components, then the radial component is rotated to its output direction (dependent on its pupil position) before the components are recomposed as a three-dimensional Jones vector. The easiest way to achieve this is to rotate the input illumination about the z-axis until it lies on the x-axis of the pupil. Then rotate the fields about the y-axis according to how far off-axis the beam is in the pupil, before rotating the resulting field back around the z-axis to its original position.

To describe the beam, we define a rotation matrix $R(\theta, \phi)$ with a polar angle θ $(0 < \theta < \pi)$ in the x-z plane and an azimuthal angle ϕ $(0 < \phi < 2\pi)$ in the x-y plane as

$$R(\theta,\phi)=\begin{pmatrix}cos\,\phi & -sin\,\phi & 0\\ sin\,\phi & cos\,\phi & 0\\ 0 & 0 & 1\end{pmatrix}\cdot\begin{pmatrix}cos\,\theta & 0 & -sin\,\theta\\ 0 & 1 & 0\\ sin\,\theta & 0 & cos\,\theta\end{pmatrix}\cdot\begin{pmatrix}cos\,\phi & sin\,\phi & 0\\ -sin\,\phi & cos\,\phi & 0\\ 0 & 0 & 1\end{pmatrix}$$
$$=\begin{pmatrix}cos\,\theta\,cos^2\phi+sin^2\phi & (cos\,\theta-1)\,cos\,\phi\,sin\,\phi & -cos\,\phi\,sin\,\theta\\ (cos\,\theta+1)\,cos\,\phi\,sin\,\phi & cos^2\phi+cos\,\theta\,sin^2\phi & -sin\,\theta\,sin\,\phi\\ cos\,\phi\,sin\,\theta & sin\,\theta\,sin\,\phi & cos\,\theta\end{pmatrix} \quad 3\text{-}1$$

, where · denotes the matrix product. The matrix rotates a 3-D Cartesian coordinate system about z axis before rotating it about y axis and then rotates it back about z axis. In a SIM set-up, the polar angle $\sin\theta$ is either $\frac{NA_{ill}}{n}$ or $\frac{-NA_{ill}}{n}$ for all beams because they are symmetric with respect to the optical axis (z), where NA_{ill} is the illumination numerical aperture. However, the azimuthal one ϕ is different at each beam, also changes with the location of the beam when the illumination pattern rotates at different time points. As illustrated in Figure 3-2, in the SIMulator we provide four SIM modes, including conventional 2 beams, hexagonal 3 beams and that at a right-angle, also conventional 2 beams for 3D-SIM.

SIM mode	θ_b	ϕ_{ba}
Conventional 2 beams	$\theta_1=\frac{NA_{ill}}{n},\theta_2=\frac{-NA_{ill}}{n}$	$\phi_{11},\phi_{21}=0$ $\phi_{12},\phi_{22}=\frac{2\pi}{3}$ $\phi_{13},\phi_{23}=\frac{4\pi}{3}$
Hexagonal	$\theta_1=\frac{NA_{ill}}{n}$	$\phi_{11}=0,\phi_{21}=\frac{2\pi}{3},\phi_{31}=\frac{4\pi}{3}$
Hexagonal at right angles	$\theta_1=\frac{NA_{ill}}{n}$	$\phi_{11}=0,\phi_{21}=\frac{\pi}{2},\phi_{31}=\pi$
Conventional 3 beams	$\theta_1=\frac{NA_{ill}}{n},\theta_2=0,\theta_3=\frac{-NA_{ill}}{n}$	$\phi_{11},\phi_{21},\phi_{31}=0$ $\phi_{12},\phi_{22},\phi_{32}=\frac{2\pi}{3}$ $\phi_{13},\phi_{23},\phi_{33}=\frac{4\pi}{3}$

Table 3-1.Angles in different SIM modes. θ for the two hexagonal modes is taken as the same value for three beams for the simplicity.

However, Figure 3-2 only shows the initial positions and the polar angles of beams especially in conventional 2-D and 3-D modes, while the change with the orientation of illumination patterns is accounted for to determine the two angles, which are shown in Table 3-1, where b and a subscript denotes the number of the beam and the angle, respectively. The polarisation states of beams also need to be taken into account because they affect the contrast seen in illumination patterns. The polarisation of each beam is calculated as the matrix product of a rotation matrix according to the polarisation direction for different beam positions in the pupil and an initial field vector describing the basic polarisation (linear or circular). As presented in Figure 3-3, five polarisation cases are simulated.

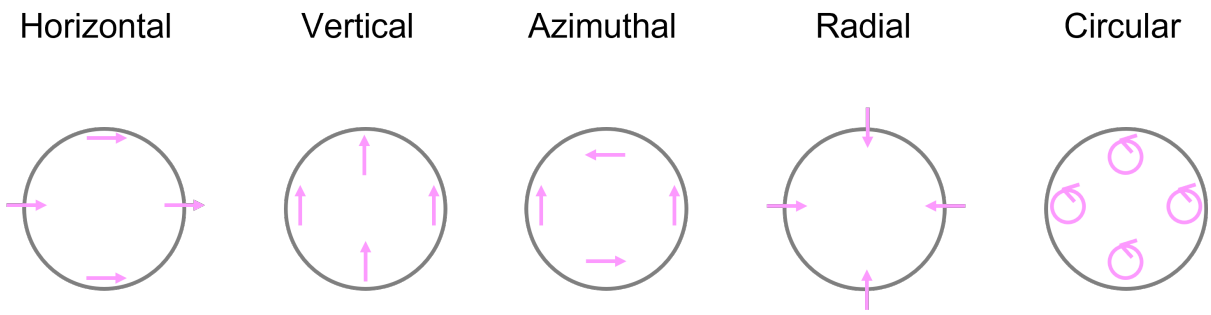


Figure 3-3. Polarisation states. The number of beams is taken as 4, which can be adjusted in the SIMulator.

For horizontal, vertical, horizontal and circularly polarised fields, the input polarisation is identical for all beams, while for azimuthal and radial cases the pupil Jones vector $\bar{S}_p$ rotates by the pupil position angle ϕ, in the x-y plane.

Polarisation	Horizontal	Vertical	Azimuthal	Radial	Circular
Pupil Jones Vector $\bar{S}_p$	$\begin{pmatrix}1 & 0\\0 & 1\\0 & 0\end{pmatrix}\cdot\begin{pmatrix}1\\0\end{pmatrix}$	$\begin{pmatrix}1 & 0\\0 & 1\\0 & 0\end{pmatrix}\cdot\begin{pmatrix}0\\1\end{pmatrix}$	$\begin{pmatrix}cos\,\phi & -sin\,\phi\\sin\,\phi & cos\,\phi\\0 & 0\end{pmatrix}\cdot\begin{pmatrix}0\\1\end{pmatrix}$	$\begin{pmatrix}cos\,\phi & -sin\,\phi\\sin\,\phi & cos\,\phi\\0 & 0\end{pmatrix}\cdot\begin{pmatrix}1\\0\end{pmatrix}$	$\begin{pmatrix}1 & 0\\0 & 1\\0 & 0\end{pmatrix}\cdot\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{i}{\sqrt{2}}\end{pmatrix}$

Table 3-2.Jones vectors of polarised beams travelling parallel to the optic axis (z-direction). The imaginary unit is denoted using j.

Table 3-2 shows the pupil Jones vector $\bar{S}_p$ for different polarisation states, where · denotes dot multiplication and i is the imaginary unit. Given the lens rotation matrix and the pupil plane Jones vector, the Jones' vector of a beam in sample space can be calculated as:

$$\bar{S}_b = \begin{pmatrix} S_{x_{ba}} \\ S_{y_{ba}} \\ S_{z_{ba}} \end{pmatrix} = R(\theta_b, \phi_{ba}) \cdot \bar{S}_{pb} \quad 3\text{-}2$$

, where b is the number of the beam and a is the number of the beam angle.

3.2.2. Shifting the phase of Jones vectors to change the illumination pattern position

In addition to the amplitude, the relative phase of the illumination beams is critical in determining how a sample is illuminated. In particular altering the relative phase of the beams is used to shift the illumination pattern in the sample space. The phase of each beam in sample space can be calculated from the dot-product of the beam's k-vector and the spatial co-ordinate in sample space. The k-vector in sample space for a beam is essentially the pupil plane k-vector, oriented in the z-direction, multiplied by the same lens rotation matrix $R(\theta_b, \phi_{ba})$. It should be noted that for each illumination pattern to reconstruct a super-resolved image, there is a phase shift for each beam φ_{ba} that shifts the pattern in sample space. Any extra phase error e_{phase} can be added here for testing the SIMulator and processing algorithms. The total phase of the field in each of the beams at a point (x, y, z) in sample space can be written as

$$\Phi_b(x, y, z) = (x \quad y \quad z) \cdot R(\theta_b, \phi_{ba}) \cdot \begin{pmatrix} 0 \\ 0 \\ k_{ill} \end{pmatrix} + \varphi_b + e_{phase} \quad 3\text{-}3$$

, and k_{ill} is the wave number of the illumination beams that is defined as $\frac{2\pi n}{\lambda_{ill}}$. Given the result from equation 3-2, the polarised total field in sample space can therefore be written as:

$$\bar{S}_{tot}(x, y, z) = \sum_b \bar{S}_{pb} e^{-j(\Phi_b(x,y,z))} \quad 3\text{-}4$$

. This field is still a Jones' vector but the different intensities of its polarisation components can be calculated from the element-by-element product of Jones' vector components field and their complex conjugate:

$$\bar{I}_{tot}(x, y, z) = \begin{pmatrix} I_x \\ I_y \\ I_z \end{pmatrix} = \bar{S}_{tot} \times \bar{S}^*_{tot} \quad 3\text{-}5$$

. These are important later as they determine how light is absorbed by oriented dipoles. In the simplest case for now we can write the total intensity as just $I_x + I_y + I_z$ in order to look at some example patterns. In the SIMulator, twenty main scenarios (4 SIM modes x 5 polarisations) can be simulated. Figure 3-4 displays some examples of illumination patterns generated by the SIMulator, in each case presenting one illumination pattern out of those required for each SIM geometry.

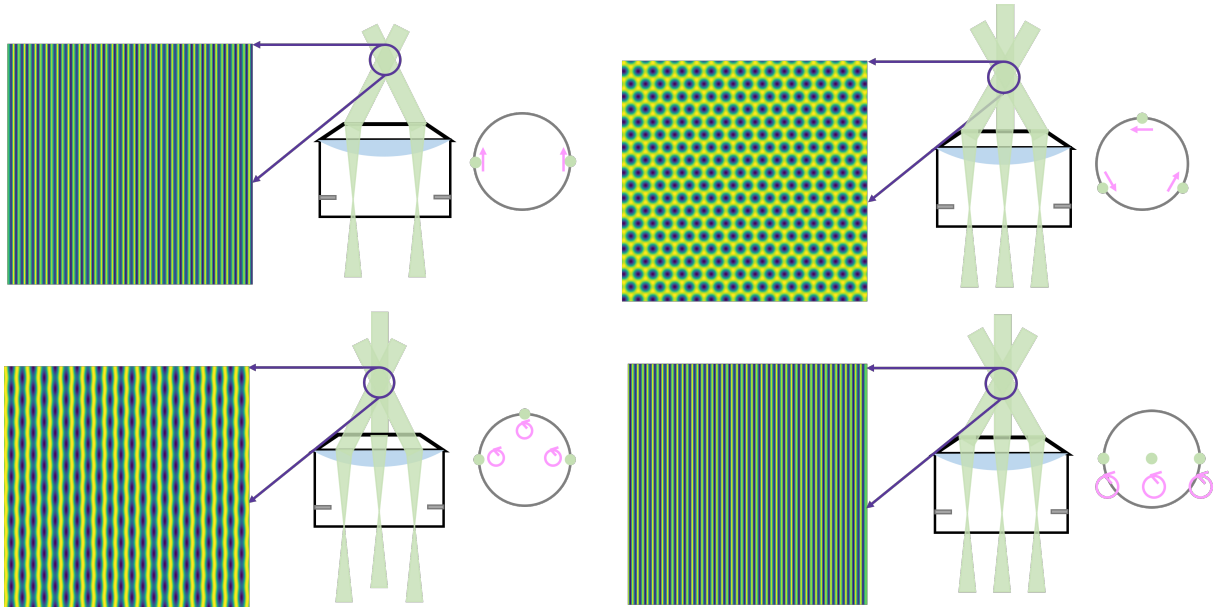


Figure 3-4. Four examples of illuminations with different polarised beams. The illumination changes with different parameters and situations in the SIMulator.

3.3. imaging from a sample space to a camera plane

For each point in the point cloud, we can determine how much it is excited by the 3-dimensional illumination intensity (as calculated in section 3.2). We also need to determine what the image of that point looks like at the camera. For this we need to know what the image of a point looks like in the detection imaging system. The response of the microscope system is analysed by studying imaging from a sample plane to a camera plane. Imaging involves capturing and analysing waves, which are described using fields. As presented in Figure 3-5, the fields locate at two sides of a lens depend on their movements with respect to the lens's focal plane.

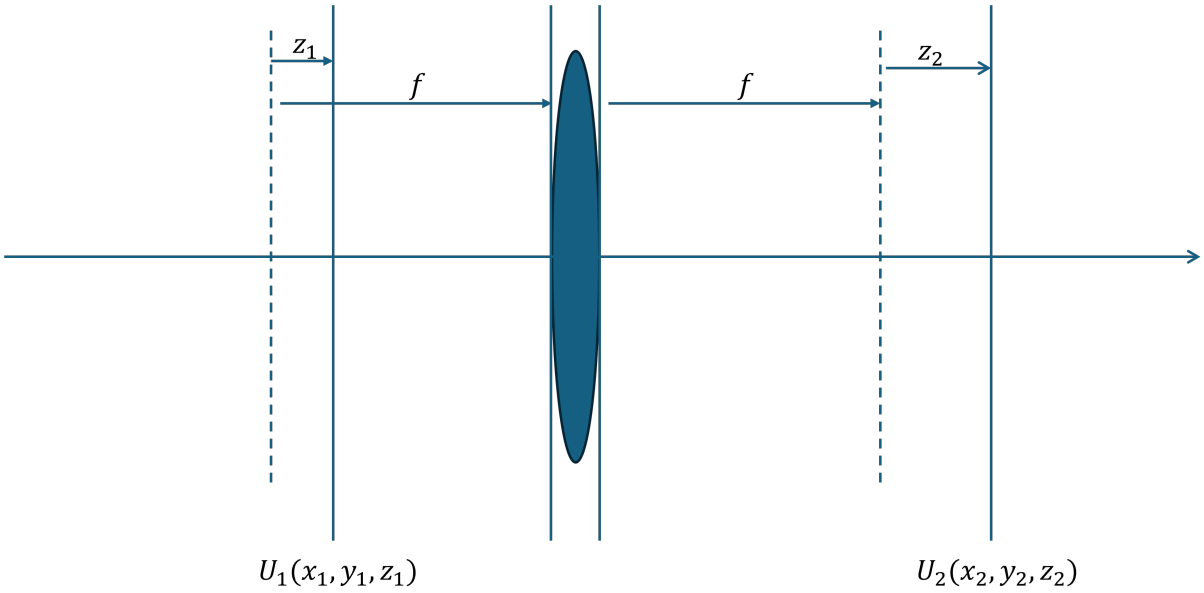


Figure 3-5. The fields close to the front and back focal plane of a lens.

The equation to calculate the field in a plane near the back focal plane of a lens $U_2(x_2, y_2, z_2)$ as a function of the field in a plane near the front focal plane of the lens $U_1(x_1, y_1, z_1)$ is

$$U_2(x_2, y_2, z_2) = \exp\left[j\frac{k_0 z_1}{2f^2}(x_2^2+y_2^2)\right]\iint_{-\infty}^{\infty} U_1(x_1, y_1, z_1)\exp\left[j\frac{k_0 z_2}{2f^2}(x_1^2+y_1^2)\right]\exp\left[-j\frac{k_0}{f}(x_1x_2+y_1y_2)\right]dx_1dx_2 \quad 3\text{-}6$$

, where $k_0 = \frac{2\pi}{\lambda}$ is the free space wave number and λ is the wavelength of light. The equation is valid only for systems with a low NA. considering a microscope system, as shown in Figure 3-6, an object is imaged by an objective lens onto a pupil plane and then the field is projected onto the camera plane using a tube lens.

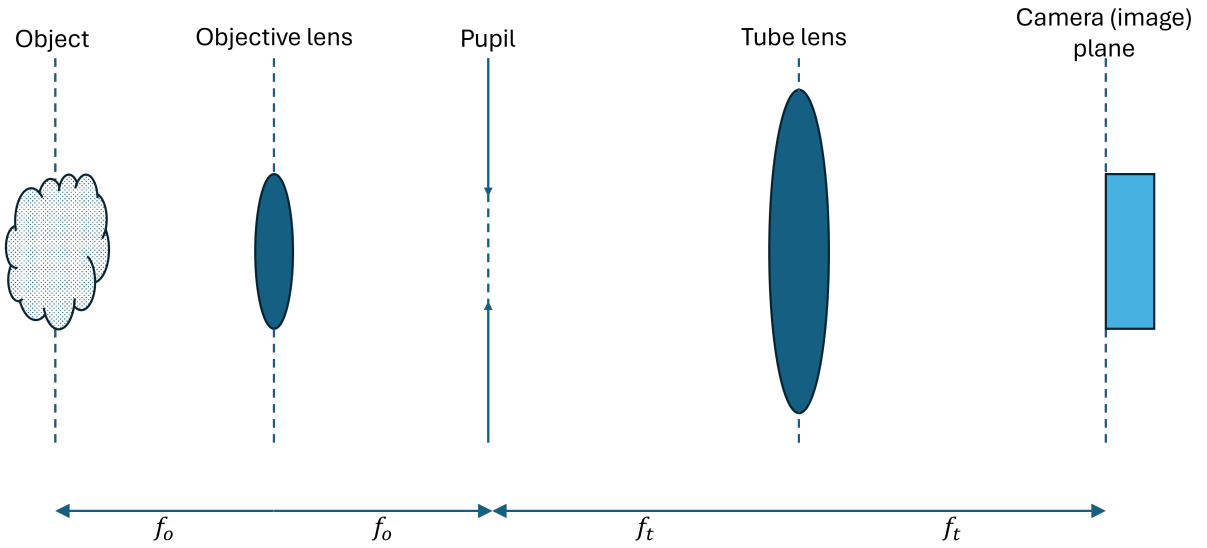


Figure 3-6. The layout of the microscope for imaging from a sample space to a camera plane.

Typically for a high magnification, M, the objective has a short focal length, f_o, and numerical aperture, NA, while the tube lens has a much longer focal length, $f_t = M \times f_o$. The two lenses share a pupil plane that is positioned in the back focal plane of the objective and the front focal plane of the tube-lens. The numerical aperture of the tube lens is thus a factor of $M\times$ smaller than that of the objective. The sample might be three-dimensional, but the camera is a two-dimensional detector in the back focal plane of the tube lens, and it detects intensity. The sample here consists of fluorescent molecules that emit light at a wavelength λ_e when excited by illumination light at a wavelength λ_i. Emission from different molecules in the sample is uncorrelated, so we can consider the expected intensity detected at the camera plane to be the sum of the expected intensities detected from each of the fluorophores independently in the sample.

3.3.1. Calculating the field in the pupil plane for a single point in the object

In order to understand the imaging process, first we should determine the image of a single fluorophore located at a position (x, y, z) in sample space, where the origin $(0,0,0)$ lies on axis in the front focal plane of the objective. The point will emit a spherical wave with wavevectors (k_x, k_y, k_z) where $k_x^2 + k_y^2 + k_z^2 = k_0^2$. In k-space the wave corresponds to a sphere, but it has a phase given by:

$$\phi_k = k_x x + k_y y + k_z z \qquad 3\text{-}7$$

. As the lens has a high numerical aperture, we need to be careful about how the lens focuses the emitted wave in its back focal plane. For high numerical aperture objective lenses, it is important to minimise aberrations, particularly coma aberrations, to ensure that the lens obeys Abbe's sine condition [60]. Here a ray travelling at an angle θ to the optic axis is focussed in the back focal plane at a height $h = f\sin\theta$, where f is the focal length of the lens in object space. It is usual that a high numerical aperture objective is also an immersion lens where the sample space has a refractive index n so we should take account of this by writing:

$$k_x^2 + k_y^2 + k_z^2 = n^2 k_0^2 \qquad \text{3-8}$$

. As the emitted spherical wave will have wave number nk_0, and $f = nf_o$. Waves with wavevector (k_x, k_y, k_z) will therefore appear in the pupil at a height (x_p, y_p) given by:

$$x_p = nf_o\frac{k_x}{nk_0} = f_o\frac{k_x}{k_0}$$
$$y_p = f_o\frac{k_y}{k_0} \qquad \text{3-9}$$

. The total wave field in the pupil will be given by:

$$U_p(x_p, y_p; x, y, z) = \exp\left(j\left(k_x x + k_y y + k_z z\right)\right) = \exp\left(j\frac{k_0}{f_o}\left(x_p x + y_p y\right)\right)\exp(jk_z z) \qquad \text{3-10}$$

Substituting $k_z^2 = n^2k_0^2 - k_x^2 - k_y^2$ (see equation 3-8) into equation 3-10, gives:

$$U_p(x_p, y_p; x, y, z) = \exp\left(j\frac{k_0}{f_o}\left(x_p x + y_p y\right)\right)\exp\left(jz\sqrt{n^2k_0^2 - k_x^2 - k_y^2}\right)$$
$$= \exp\left(j\frac{k_0}{f_o}\left(x_p x + y_p y\right)\right)\exp\left(jznk_0\sqrt{1 - \frac{k_x^2}{n^2k_0^2} - \frac{k_y^2}{n^2k_0^2}}\right) \qquad \text{3-11}$$
$$= \exp\left(j\frac{k_0}{f_o}\left(x_p x + y_p y\right)\right)\exp\left(jznk_0\sqrt{1 - \frac{x_p^2}{n^2f_o^2} - \frac{y_p^2}{n^2f_o^2}}\right)$$

. Note that the final phase term corresponds to the defocus of the point in sample space which for low NA systems where $x_p^2 + y_p^2 \ll n^2f_o^2$ is usually approximated to

$$U_p(x_p, y_p; x, y, z) \approx \exp\left(j\frac{k_0}{f_o}(x_p x + y_p y)\right)\exp\left(jznk_0\left(1 - \frac{x_p^2}{2n^2 f_o^2} - \frac{y_p^2}{2n^2 f_o^2}\right)\right)$$
$$= \exp\left(j\frac{k_0}{f_o}(x_p x + y_p y)\right)\exp(jznk_0)\exp\left(-j\frac{zk_0}{2nf_o^2}(x_p^2 + y_p^2)\right) \quad 3\text{-}12$$

And ignoring constant phase terms, the pupil's total field is written as follows:

$$U_p(x_p, y_p; x, y, z) \approx \exp\left(j\frac{k_0}{f_o}(x_p x + y_p y)\right)\exp\left(-j\frac{zk_0}{2nf_o^2}(x_p^2 + y_p^2)\right) \quad 3\text{-}13$$

3.3.2. Calculating the image of the point on the camera

For the tube lens we just need to perform a simple Fourier transform as we are calculating the field in the camera plane (back focal plane of the tube-lens) as a function of the field in the pupil plane (front focal plane of the tube-lens) as:

$$U_c(x_c, y_c; x, y, z) = \iint_{pupil} \exp\left(-j\frac{k_0}{f_t}(x_p x_c + y_p y_c)\right) U_p(x_p, y_p; x, y, z) dx_p dy_p \quad 3\text{-}14$$

. Substituting equation 3-12 in equation 3-14, the field in the camera plane is

$$U_c(x_c, y_c; x, y, z) \quad 3\text{-}15$$
$$= \iint_{pupil} \exp\left(-j\frac{k_0}{f_t}(x_p x_c + y_p y_c)\right)\exp\left(j\frac{k_0}{f_o}(x_p x + y_p y)\right)\exp\left(jznk_0\sqrt{1 - \frac{x_p^2}{n^2 f_o^2} - \frac{y_p^2}{n^2 f_o^2}}\right) dx_p dy_p$$

. Note that the image of the point simply shifts laterally on the camera as the point shifts laterally in the sample as we can write:

$$U_c(x_c, y_c; x, y, z) = U_c\left(x_c + \frac{f_t}{f_o}x, y_c + \frac{f_t}{f_o}y; 0,0, z\right) = U_c(x_c + Mx, y_c + My; 0,0, z) \quad 3\text{-}16$$

. In other words, the image is simply a magnified (though inverted) image of the sample, though blurred by a z-dependent point spread function. We therefore only need to calculate the image of an on-axis point at $(x, y) = (0,0)$ to know the image of a point shifted to any other lateral position. Remembering that we are interested in the intensity at the sample plane, this is given by:

$$I_c(x_c, y_c; 0,0, z) \tag{3-17}$$

$$= \left| \iint_{pupil} \exp\left(-j\frac{k_0}{f_t}\left(x_p x_c + y_p y_c\right)\right) \exp\left(jznk_0\sqrt{1-\frac{x_p^2}{n^2 f_o^2}-\frac{y_p^2}{n^2 f_o^2}}\right) dx_p dy_p \right|^2$$

3.3.3. Calculating the image of a point cloud

Equation 3-17 describes a three-dimensional point spread function, that we can calculate across a grided three-dimensional array of values in (x_c, y_c, z). However, we would like to place the points in the point cloud at arbitrary positions in the sample, (x_s, y_s, z_s). To shift the point spread function by an arbitrary amount we can use the Fourier space again and the Fourier shift theorem. First, we calculate the 3-dimensional Fourier transform of the 3-D PSF, which give the 3-D optical transfer function (OTF).

$$OTF(\kappa_x, \kappa_y, \kappa_z) = \iiint_{PSF} I(x_c, y_c, z) \exp\left(-j\left(\kappa_x x_c + \kappa_y y_{c+} \kappa_z z\right)\right) dx_c dy_c dz \tag{3-18}$$

We then multiply OTF by a tilted phase and inverse Fourier transform the result to give the shifted PSF.

$$I_c(x_c, y_c; x_s, y_s, z_s) \tag{3-19}$$

$$= \iiint_{\kappa-space} OTF(\kappa_x, \kappa_y, \kappa_z) \exp\left(j\left(\kappa_x x_c + \kappa_y y_{c+} \kappa_z 0\right)\right) \exp\left(j\left(\kappa_x M x_s + \kappa_y M y_{s+} \kappa_z z_s\right)\right) d\kappa_x d\kappa_y d\kappa_z$$

$$= I_c(x_c + M x_s, y_c + M y_s; z_s)$$

Remembering that we only need calculate the output PSF at the camera plane (effectively at $z_c = 0$). For a complete point cloud with N individual points we just sum over the individual shifted intensity PSFs, which in turn can calculated as an inverse Fourier transform over the OTF multiplied by the summation of the individual phase tilt functions:

$$\sum_{x_s, y_s, z_s} I_c(x_c, y_c; x_s, y_s, z_s) \tag{3-20}$$

$$= \iiint_{\kappa-space} OTF(\kappa_x, \kappa_y, \kappa_z) \exp\left(-j\left(\kappa_x x_c + \kappa_y y_{c+} \kappa_z 0\right)\right) \sum_{x_s, y_s, z_s} \exp\left(-j\left(\kappa_x M x_s + \kappa_y M y_{s+} \kappa_z z_s\right)\right) d\kappa_x d\kappa_y d\kappa_z$$

3.4. Vectorial calculation to take account of dipole emitters

However, in reality, the PSFs and OTFs are affected by the molecular nature of the individual fluorophores that are not isotropic emitters as they will both absorb and re-emit light as dipoles. The light that they emit will be both polarised and have a dipole emission pattern dependent on their orientation. In addition, the polarisation of the emitted light will change as the rays refract through the lens in a similar way to was described for the illumination beams.

Each individual fluorescent molecule may be considered a dipole absorber(emitter). Therefore, as shown in Figure 3-7 (a), both the absorption and emission of each fluorescent molecule can be described by dipole radiation, which is anisotropic. For instance, as seen in Figure 3-7 (b) and (c), when the direction we look the dipole from is along z, the radiation field appears radial and has a dark centre, while as the direction changes looking at the dipole from y, the field is largely linear and bright in the centre. Hence, both the polarisation of the excitation light and directions of the emission dipoles can affect the resultant simulated PSFs. To account for this, the electric field is described by a 3-D complex vector.

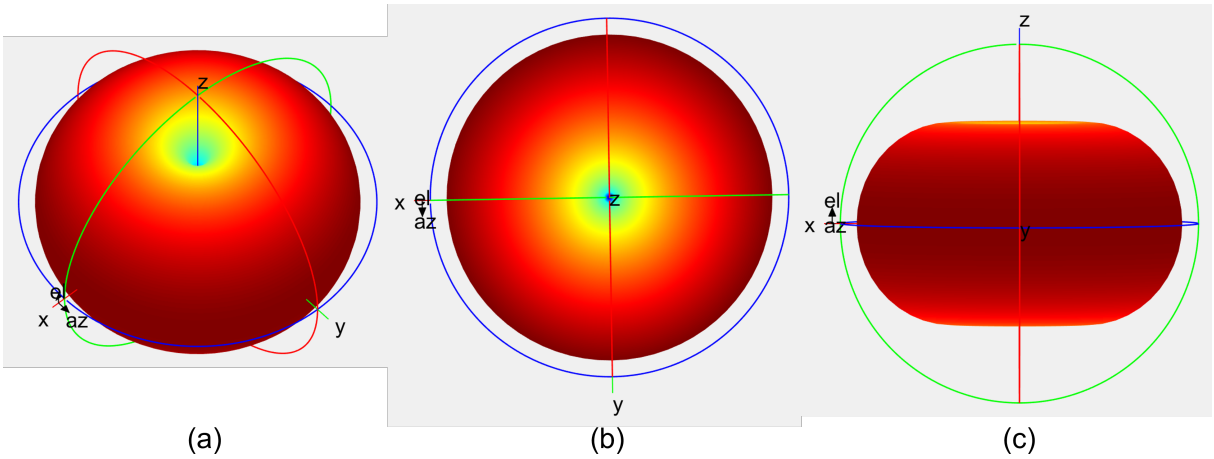


Figure 3-7. (a) A dipole radiation at a random direction. (b) Dipole radiation at x-y plane and x-z plane (c). the radiation changes with the observation direction of the objective in a system.

The vectorial nature of the field in a pupil plane has been a subject of extensive study by many researchers over the years, dating back to approximately 105 years ago [61][62]. A notable theory is the Richards and Wolf theory, describing the electric and magnetic fields and extended it to include phase and polarisation in the pupil plane [63]. However, in those scenarios the pupil field is deduced according to the polarised illuminating light, which features an opposite direction to our case (looking at the pupil from the dipole). In our scenario, the complex amplitude of the pupil function p is influenced not only by the polarisation of the

incident light, but also by the orientation of the dipole emitters. Based on [61], [62] and [63], As discussed by Clegg in his PhD thesis [64], when imaging a dipole, the amplitude of the fields in the pupil should be modified by the following factors:

$$\frac{1}{\sqrt{\cos\theta}}\begin{pmatrix}\cos^2\phi\cos\theta+\sin^2\phi & (\cos\theta-1)\sin\phi\cos\phi & -\sin\theta\cos\phi\\ (\cos\theta-1)\sin\phi\cos\phi & \sin^2\phi\cos\theta+\cos^2\phi & -\sin\theta\sin\phi\end{pmatrix} = \begin{pmatrix}A_{xx} & A_{xy} & A_{xz}\\ A_{yx} & A_{yy} & A_{yz}\end{pmatrix} \quad 3\text{-}21$$

, where $\cos\theta=\frac{k_z}{nk_0}$, $\sin\theta=\frac{k_r}{nk_0}$, $\cos\phi=\frac{k_x}{k_r}$, $\sin\phi=\frac{k_x}{k_r}$, and the apodisation factor A_{ab} (a,b can be x,y,z) is the amplitude of the component of the field in the a-direction in the pupil when the dipole is oriented in the b-direction. Hence, the apodisation factor can be written as

$$A_{xx}=\sqrt{\frac{nk_0}{k_z}}\frac{nk_0{k_y}^2+k_z{k_x}^2}{nk_0{k_r}^2} \quad A_{xy}=\sqrt{\frac{nk_0}{k_z}}\frac{k_xk_y(k_z-nk_0)}{nk_0{k_r}^2} \quad A_{xz}=\sqrt{\frac{nk_0}{k_z}}\frac{k_x}{nk_0} \quad 3\text{-}22$$
$$A_{yx}=\sqrt{\frac{nk_0}{k_z}}\frac{k_xk_y(k_z-nk_0)}{nk_0{k_r}^2} \quad A_{yy}=\sqrt{\frac{nk_0}{k_z}}\frac{nk_0{k_x}^2+k_z{k_y}^2}{nk_0{k_r}^2} \quad A_{yz}=\sqrt{\frac{nk_0}{k_z}}\frac{k_y}{nk_0}$$

. Note that while the dipole can be oriented in any of the x, y or z directions in the sample, it can only be oriented in the (x,y) plane in the pupil and the camera plane as these are low NA and paraxial. Expressing the terms A_{ab} in terms of k_x and k_y is convenient as these can be calculated as arrays when using FFT routines to perform the calculations. In fact, although the light in our case emits from the dipole (sample) to the pupil plane, these apodisation factors are the same as that in Boruah's paper [65] where the polarised light passing through the pupil plane is focused onto the sample. Namely, along the two reciprocal proportions, the field components of the pupil plane can be described as the same. To account for the optical aberration in the detective path, Zernike circle polynomials is used. $Zernike$ in the equation represents the sum of Zernike polynomials. For example, the polynomial of the spherical aberration can be written as

$$spherical=\sqrt{5}\beta\left[6\times\left(\frac{k_r}{k_{r\,max}}\right)^4-6\left(\frac{k_r}{k_{r\,max}}\right)^2+1\right] \quad 3\text{-}23$$

, where coefficient β indicates the magnitude of the spherical aberration.

Given the phase of the field as $e^{j\left(z\sqrt{{k_0}^2-{p_A(0)}^2}+Zernike\right)}$, the amplitude of PSF E in each z plane corresponding to the six cases above can be written as

$$E_{ab} = iFFT\left[A_{ab}e^{j\left(z\sqrt{{k_0}^2-{p_A(0)}^2}+Zernike\right)}\right], \forall a: x, y, \forall b: x, y, z \quad \text{3-24}$$

. The six components E_{ab} can be grouped and rewritten as two vectors, in terms of the x and y field components:

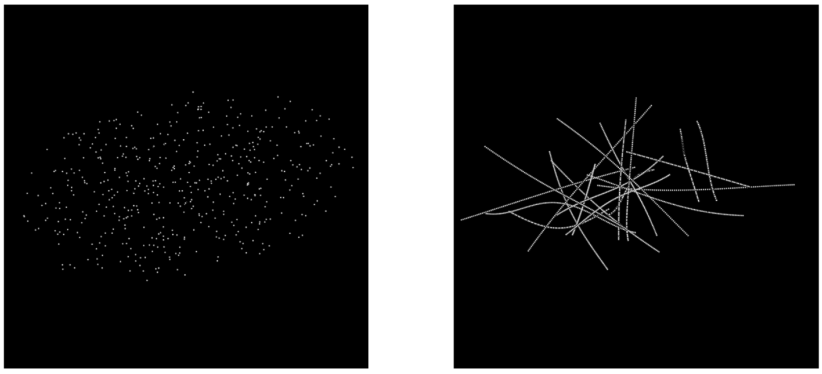
$$\underline{E_x} = \left(E_{xx}, E_{xy}, E_{xz}\right), \qquad \underline{E_y} = \left(E_{yx}, E_{yy}, E_{yz}\right) \quad \text{3-25}$$

3.5. Computing the simulation of SIM imaging

The development in modern computers facilitates fast calculation of simulating SIM imaging. Initially, a point-cloud sample can be simulated. The computing of the illumination calculation is omitted in this section since there are mainly dots multiplications between matrices. In terms of the emission from the sample, according to the theory explained above, scalar and vectorial PSFs and OTFs in the pseudo system can be computed using the Fast Fourier transform (FFT). In vectorial cases, the dipole orientations are considered. Finally, 2-D raw SIM images can be simulated, and the speed of the most time-consuming part (phase tilt calculation) is tested in a few of scenarios with different accelerations.

3.5.1. Simulating objects as point clouds

Given the optical system, the simulation requires samples to be ‘prepared’ for virtual ‘imaging’, i.e., modelled. Here, the samples being illuminated in SIM systems are objects simulated as point clouds in sample space. As shown in Figure 3-8, in the SIMulator we can generate two types of 3-D samples: spheroidal and filament- point clouds, mimicking common biological fluorophore distributions observe in cells.



(a) (b)

Figure 3-8. A slice of 3-D samples: Spheroidal (a) and filament (b) point-clouds.

When generating a spheroidal point-cloud distribution, the number of points n_{points}, the radius at the z=0 plane r (given that the longitudinal axis of the point-cloud spheroid is along the z axis) and the depth dep (along z direction) can be adjusted by users. Initially, 3-D coordinate values of some random points are generated within a range of a cube with radius as r:

$$\begin{aligned} -r \le px_n < r \\ -r \le py_n < r \\ -dep \le pz_n < dep, n = 1,2,3 \dots n_{poinnts} \end{aligned} \quad 3\text{-}26$$

, where n is the number of points. Taking (0,0,0) as the origin, we can then calculate the relative position of each point and then select a spheroid from the cube by choosing the points inside the spheroid:

$$\sqrt{px_n{}^2 + py_n{}^2 + pz_n{}^2} < r \quad 3\text{-}27$$

. When generating a filament-point-cloud object, we can control the number of filaments n_f, the length L and the fineness (spacing between points) dL of the filaments. The centres of each filament are chosen in the $x - y$ plane as a random coordinate as

$$(x_c, y_c, 0), -\frac{L}{2} < x_c \le \frac{L}{2}, -\frac{L}{2} < y_c \le \frac{L}{2} \quad 3\text{-}28$$

. An azimuthal angle (an angle in the x-y plane) α $(0 \le \alpha < \pi)$ and a polar angle β $\left(-\frac{\pi}{4} \le \beta < \frac{\pi}{4}\right)$ are chosen to describe the direction of a filament. In the simulation, the angles are chosen as random values in their range, respectively. Using the centre and angles, the starting and ending points: (x_s, y_s, z_s) and (x_e, y_e, z_e) of a filament can be calculated as follows:

$$\begin{aligned} x_s = x_c + \frac{l}{2} * \cos(\alpha) * \cos(\beta), y_s = y_c + \frac{l}{2} * \sin(\alpha) * \cos(\beta), z_s = z_c + \frac{l}{2} * sin\,(\beta) \\ x_e = x_c - \frac{l}{2} * \cos(\alpha) * \cos(\beta), y_e = x_c - \frac{l}{2} * \sin(\alpha) * \cos(\beta), z_e = z_c - \frac{l}{2} * sin\,(\beta) \end{aligned} \quad 3\text{-}29$$

. In order to simulate more realistic filaments akin to those found in biological images, a perturbation is added in the simulation: a random sinusoidal function is added on the straight filament to make it wavy. Using random values given by $rand_n$, these perturbation factors are determined as

$$p_x = 2\pi * rand_1, f_x = \pi * rand_2, a_x = \frac{L * rand_3}{10},$$
$$p_y = 2\pi * rand_4, f_y = \pi * rand_5, a_y = \frac{L * rand_6}{10}, \quad 3\text{-}30$$
$$p_z = 2\pi * rand_7, f_z = \pi * rand_8, a_z = \frac{L * rand_9}{10}$$

, where $rand_n$ is chosen in $[0:1]$. The filaments consist of a series of points spaced by dL, so the resultant coordinates of the points in a single filament are

$$l_p = p * \frac{dL}{L}, p = 0,1,2,3, \dots \frac{L}{dL} - 1$$
$$x_p = x_s + (x_e - x_s) * l_p + a_x * cos(f_x + l_p + p_x)\ , p = 0,1,2,3, \dots \frac{L}{dL} - 1$$
$$y_p = y_s + (y_e - y_s) * l_p + a_y * cos(f_y + l_p + p_y)\ , p = 0,1,2,3, \dots \frac{L}{dL} - 1 \quad 3\text{-}31$$
$$z_p = z_s + (z_e - z_s) * l_p + a_z * cos(f_z + l_p + p_z)\ , p = 0,1,2,3, \dots \frac{L}{dL} - 1$$

, with p annotating the number of points in one filament. All points are then saved as a filament, and a number of filaments are generated as a sample, which is user definable in the SIMulator.

3.5.2. Computing scalar PSFs and OTFs

We begin with simulating an ideal optical system where the PSF and the OTF are described as complex scalar potentials. In this case, the polarisation state of each emitter in the sample is regarded to be isotropic, which is an idealised approximation. According to equation 3-17, the z-dependent PSF can be calculated as the Fourier transform of the pupil field that features a phase term. For convenience, we consider the pupil plane (k_x, k_y, k_z) in the frequency domain and the camera plane (x, y, z) in the spatial domain; thus, the inverse Fourier transform is used.

By exploiting the inverse Fast Fourier transform (iFFT), the intensity PSF at $z = 0$ plane can be defined as

$$psf_{intensity}(z) = \left| iFFT\left[p_A e^{j\left(z\sqrt{{k_0}^2 - (k_x^2 + k_y^2)} + Zernike\right)} \right] \right|^2 \quad 3\text{-}32$$

. The amplitude p_A is defined over a circle of radius of maximum lateral frequency ($k_{r_{max}}$) in the pupil plane and k_0 is the wave number of the emission light. Substituting each z value based on the depth of the emitter, the intensity $psf_{intensity}$ is calculated. To normalise the total power of the PSF as unity at the focal plane, the intensity is divided by the summation of the square of the amplitude pupil functions, and the scalar PSF is given:

$$psf_{scalar} = \frac{psf_{intensity}}{\sum {p_A}^2} \quad 3\text{-}33$$

The OTF can be calculated as the Fourier transform of the intensity PSF as

$$otf_{scalar} = FFT[psf_{scalar}] \quad 3\text{-}34$$

3.5.3. Modelling dipole orientations in the sample

Molecular fluorophores can have many properties in a sample. They could be fixed in their orientation, or they could be randomly oriented. In addition, they could be mobile and re-orientate on a timescale depending on their molecular size. They could also have a different absorption dipole orientation to their emission dipole orientation. Here we model two simple cases, which both have a completely random distribution of orientation. The difference between the two cases is that in one the dipole is considered to have a fixed orientation between excitation and emission. In the other the molecule is assumed to rotate to a completely new orientation between excitation and emission. The first case might be suitable for a large molecule such as a fluorescent protein, while the second might be suitable for a smaller molecular dye such as fluorescein.

In both cases we therefore assume an isotropic distribution of fluorophore orientation for excitation. In reality the image of a point-source in the cloud is formed from many absorption or emission cycles so this is a reasonable assumption to make if the fluorophore is not absolutely fixed. Then for the first case the preferential absorption results in a not-isotropic distribution for emission, and it is partially polarised in the pupil, while the second case the re-orientation results in an isotropic orientation again for emission. In order to model an isotropic distribution, we need to consider a set of dipole orientations that has no bias in direction. To achieve this in the code, as shown in Figure 3-9, we model the random emitter orientations of the fluorophores as a set of molecules placed along vectors to the vertices of a dodecahedron. This means that we only need to consider 10 different molecular orientations (as they are degenerate in pairs by symmetry) and that they are evenly spaced out over a sphere.

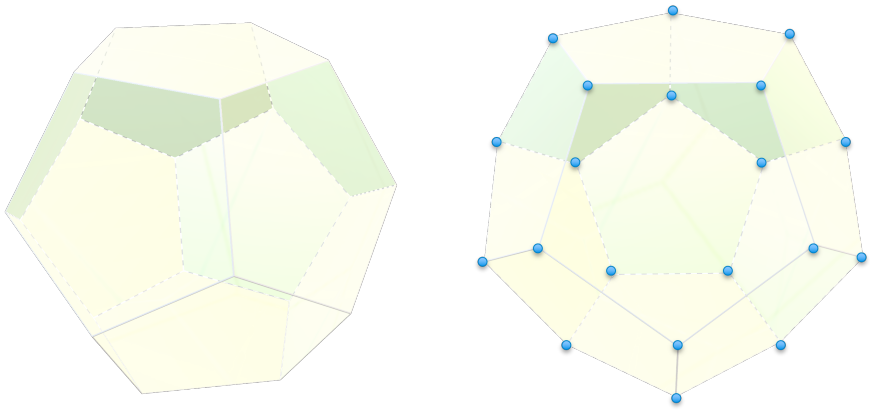


Figure 3-9. Dodecahedron and directions of single emitters (blue dots). The distribution Over the dodecahedron is isotropic. The left dodecahedron is randomly rotated compared to the right one.

In the dodecahedron model, the twenty vertices are expressed by Cartesian coordinates as:

$$\underline{V_n} = (x_n, y_n, z_n), \qquad n: 0,1,2, \dots 19 \tag{3-35}$$

, where n denotes the number of vertices. In the Python code, the values of the coordinates are normalised by dividing the square root of the summation of squared values to make the radius of the dodecahedron equal to unity, and only one half of the orientations (vertex values) are selected as the emission of all dipoles is symmetric.

3.5.4. Simulating vectorial PSFs and OTFs

Considering the dipole orientations, vectorial PSFs and OTFs are simulated in the fixed and flexible cases. Given the phase of the field as $e^{j\left(z\sqrt{{k_0}^2-{p_A(0)}^2}+Zernike\right)}$, the field of PSF E in each z plane can be written as

$$E_{ab} = iFFT\left[A_{ab}e^{j\left(z\sqrt{{k_0}^2-{p_A(0)}^2}+Zernike\right)}\right], \forall a: x, y, \forall b: x, y, z \tag{3-36}$$

, where A_{ab} is defined in equation 3-22. The six components E_{ab} can be grouped and rewritten as two vectors, in terms of the x and y field components:

$$\underline{E_x} = \left(E_{xx}, E_{xy}, E_{xz}\right), \qquad \underline{E_y} = (E_{yx}, E_{yy}, E_{yz}) \tag{3-37}$$

. To calculate the intensity from the amplitude, the structure of the point source must be considered. The total intensity I_t of the field emitted by each dipole can be calculated as the sum of x- and y- polarised components:

$$|\underline{V_n} \cdot \underline{E_x}|^2 + |\underline{V_n} \cdot \underline{E_y}|^2 \quad \text{3-38}$$

, where · represents the dot product. Depending on the polarisation of the illumination, a dipole may be excited unevenly, this can be expressed as an excitation factor α, which can be calculated as

$$\alpha_{xn} = ([1,0,0] \cdot \left[\underline{V_n}\right])^2$$

$$\alpha_{yn} = ([0,1,0] \cdot \left[\underline{V_n}\right])^2 \quad \text{3-39}$$

$$\alpha_{zn} = ([0,0,1] \cdot \left[\underline{V_n}\right])^2, \qquad n: 0,1,2, \dots 19$$

, where the first subscripts indicate the orientation of the polarised illumination. The entire intensity can thus be decomposed into three terms as I_x , I_y and I_z:

$$I_b = \sum_{n=0}^{19} \alpha_b \left(\left| \underline{V_n} \cdot \underline{E_x} \right|^2 + \left| \underline{V_n} \cdot \underline{E_y} \right|^2 \right), \forall b: x, y, z \quad \text{3-40}$$

. The intensity components can be normalised and divided by the number of dipoles to obtain the PSF x, y and z components as:

$$psf_b = \frac{I_b}{\sum p_A^2} * \frac{Nz}{Nz_n} * \frac{1}{20}, \qquad \forall b: x, y, z \quad \text{3-41}$$

When the orientation of the dipoles is flexible, the unnormalised PSF is taken as the sum of psf_x, psf_y and psf_z because the portions of these components are equally probable in this case:

$$psf_{flexible} = \frac{psf_x + psf_y + psf_z}{\sum \left(psf_x + psf_y + psf_z\right)_z} \quad \text{3-42}$$

, where the normalisation value is the sum of values over the xy-planes (z slices). Actually, the total power should be the same in each plane and this is

3-43

confirmed as a test of the correct PSF simulation. The OTF is simulated as

$$otf_{flexible} = FFT\left[psf_{flexible}\right]$$

However, if the direction of the polarised dipoles remains rigid (e.g. some fluorescent proteins), the normalisation value $norm$ is calculated as the maximum of the sum of PSF_x in $x-y$ planes as

$$norm = max\left[\sum(psf_x)_z\right] \quad 3\text{-}44$$

The three PSF components are all normalised by dividing by this value.

$$psf_{rigid_b} = \frac{psf_b}{norm}, \qquad \forall b: x, y, z \quad 3\text{-}45$$

We use normalised PSF_z as the displayed PSF for the SIMulator since the z component is more different than others. For the rigid dipoles, the OTF along three axes are calculated as three components as

$$otf_b = FFT\left[psf_{rigid_b}\right], \qquad \forall b: x, y, z \quad 3\text{-}46$$

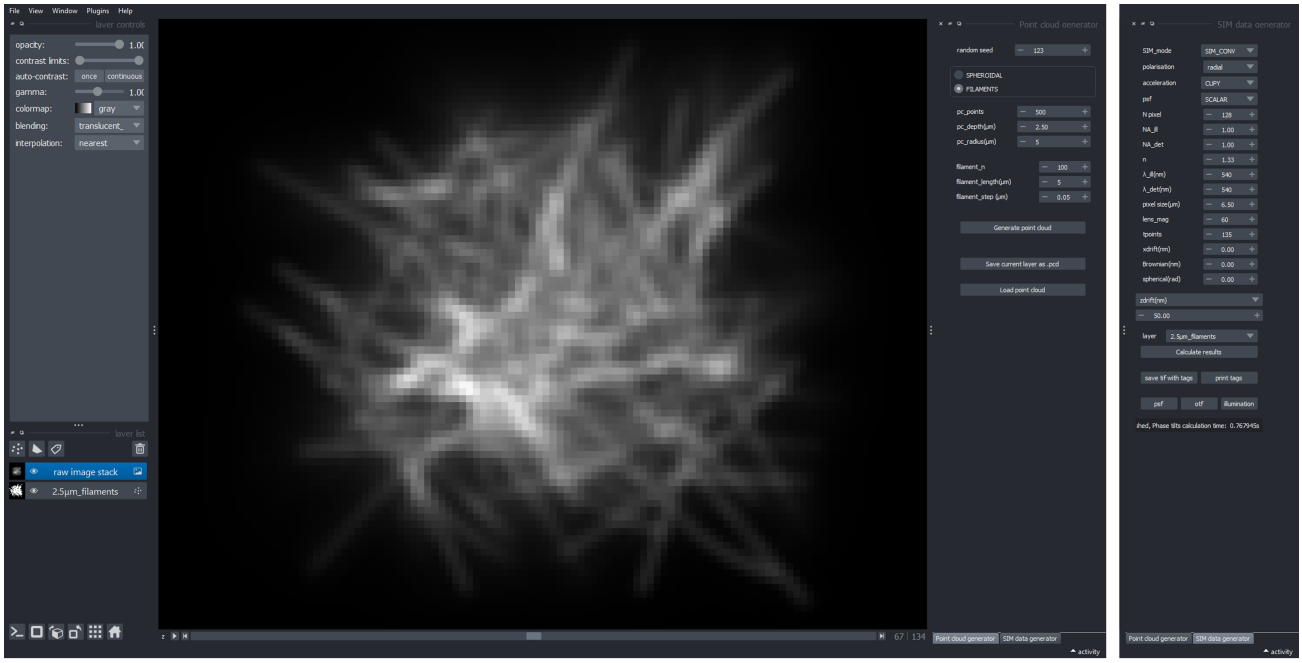
3.6. Implementation as a napari plugin: napari-generic-SIMulator

Due to the complex set-up, sensitive alignment and complicated reconstruction of SIM, it is convenient to simulate raw images in different scenarios, which can test the contrast in illumination patterns in advance and provide a ground truth for evaluating reconstruction algorithms. To this end, a plugin: napari-generic-SIMulator is written to do generic SIM simulations, and is based on napari, a Python based system for viewing, annotating and analysing multi-dimensional images.

3.6.1. Introduction of napari-generic-SIMulator

Figure 3-10 shows the GUI of napari-generic-SIMulator consisting of two widgets; a point-cloud generator and a SIM data generator. These two widgets implement a SIM experiment virtually by preparing a specimen and illuminating it with desired patterns during the acquisition of a set of images. All the parameters of the virtual experiment can be controlled and adjusted by the GUI provided by the plugin, instead of purchasing and setting up expensive lasers, SLM,

objectives and so on. Primarily, the napari-generic-SIMulator was used to assist in the development of a separate napari-SIM-processor plugin that takes the simulated images and reconstructs an enhanced resolution image of the object.



(a) (b)

Figure 3-10. GUI of napari-generic-SIMulator. (a) Point-cloud generator widget in a napari viewing window. (b) SIM data generator widget. The two widgets are usually tabified.

However, rather than the difficult real implementation of a SIM acquisition, the complexity in the experiment is provided by the calculations in the simulation. As illustrated in Figure 3-11, raw SIM images are obtained by calculations both in real and the Fourier space. In one aspect, a point-cloud with any desired motion, such as drift or Brownian motion of the sample is firstly produced. This is then combined with the information from the illumination pattern before calculating the tilted phase functions in the Fourier space. While in another aspect of the simulation, the 3-D PSF of the imaging system is calculated and then its 3-D OTF in Fourier space. The 3-D OTF is multiplied by the phase and an inverse Fourier transform is applied to obtain the raw camera images. To obtain a complete set of images all the calculations are repeated for each raw frame except for the initial simulation of the sample and the imaging system PSF.

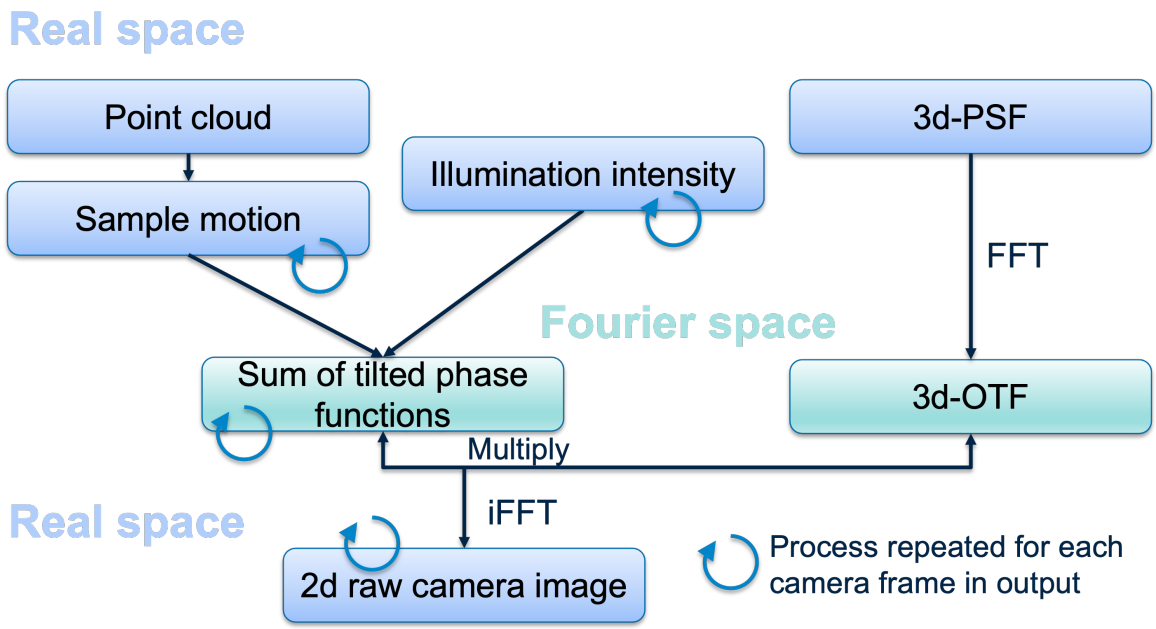


Figure 3-11. Generation of a SIM raw stack in SIMulator. Steps with the blue rotating arrow need to be repeated.

In the Python code, a 3-D point-cloud is produced as a napari 'Points' layer, and can be save as a file with '.pcd' format and loaded for later use. When a 'Points' layer is selected as the sample to be 'illuminated', the depth of the point-cloud z_{range} is calculated as the difference between the highest and lowest points, which is used to define the range of the focus of the PSF in z direction. The PSFs and OTFs can be hence simulated in different scenarios depending on the object size. Here it is also important to define the sampling coordinates in both the pupil plane and the Fourier (PSF) plane to satisfy Nyquist while minimising the processed data size and hence time. This is discussed further in Appendix: Nyquist sampling in the pupil plane and the PSF plane. The value of the 3-D OTF is calculated by taking the 3-dimensional FFT of the PSF. The structured illumination patterns are then calculated using Jones vectors as described in section 3.2 and the illumination intensities are applied to the point cloud, including any the sample motion in order to calculate the phase tilt function for each point in Fourier space. The phase tilts are then summed up over all points in the point cloud. The sum of phase tilts is multiplied with the OTF and the result is processed using an inverse Fourier transform to simulate the raw image.

Some slices of raw image stacks generated by napari-generic-SIMulator are shown below as examples.

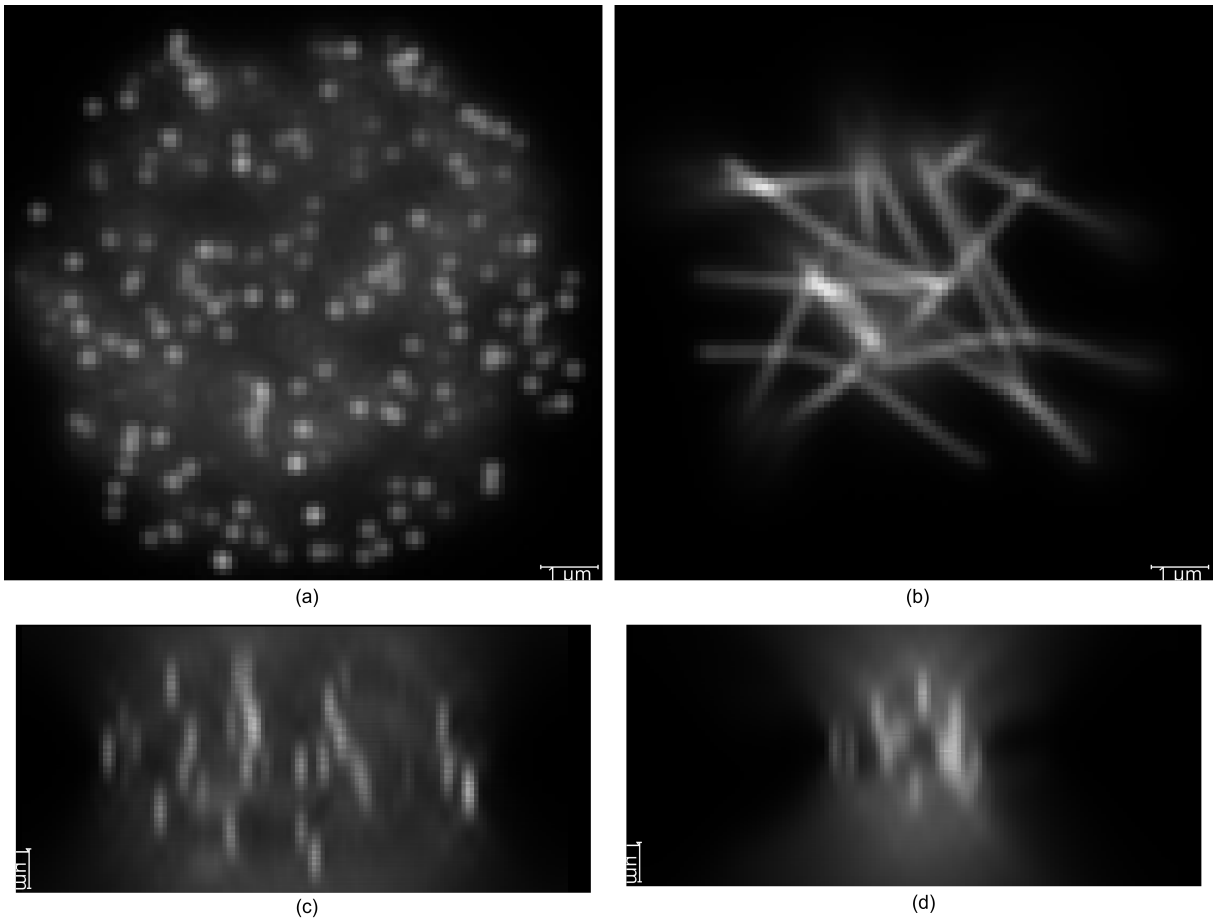


Figure 3-12. Two examples of simulated images: a spheroidal point-cloud in x-y plane (a) and x-z plane (c); a filament-point-cloud in x-y (b) and x-z plane (d).

Figure 3-12 demonstrated simulated camera images of a spheroidal and a filament-point-cloud in two planes, where the vertical lines consisted with many pixels in (c) and (d) illustrate the drifting of the sample in z direction. Without the drifting, each vertical line would be shown as a pixel. All parameters are shown in Figure 3-10. In the simulations, the two samples drift with 50 nm at each frame along the z-axis, starting from the original plane. Compared to conventional cases where frames for reconstructing one sim image keep the same z-position, each frame, in this case, is introduced with a z-drifting. The principles and the SIM reconstruction of the drifting case are explained in [38]. To compare the ground truth and the raw images obtained by the simulated imaging system, the point-cloud (green) and the raw camera image are superposed in the SIMulator, as shown in Figure 3-13. As displayed in x-y plane (a) and x-z plane (c), respectively, the spheroidal sample is superimposed with the point-cloud. (b) and (d) demonstrate the superpositions of the filaments sample and the raw images in the two planes. It can be seen that all simulated raw SIM images have the characteristics of camera images obtained in real experiments. All images showed here are at the middle of

the stack in x-y and x-z directions, and all PSFs are simulated using the scalar theory for simplicity.

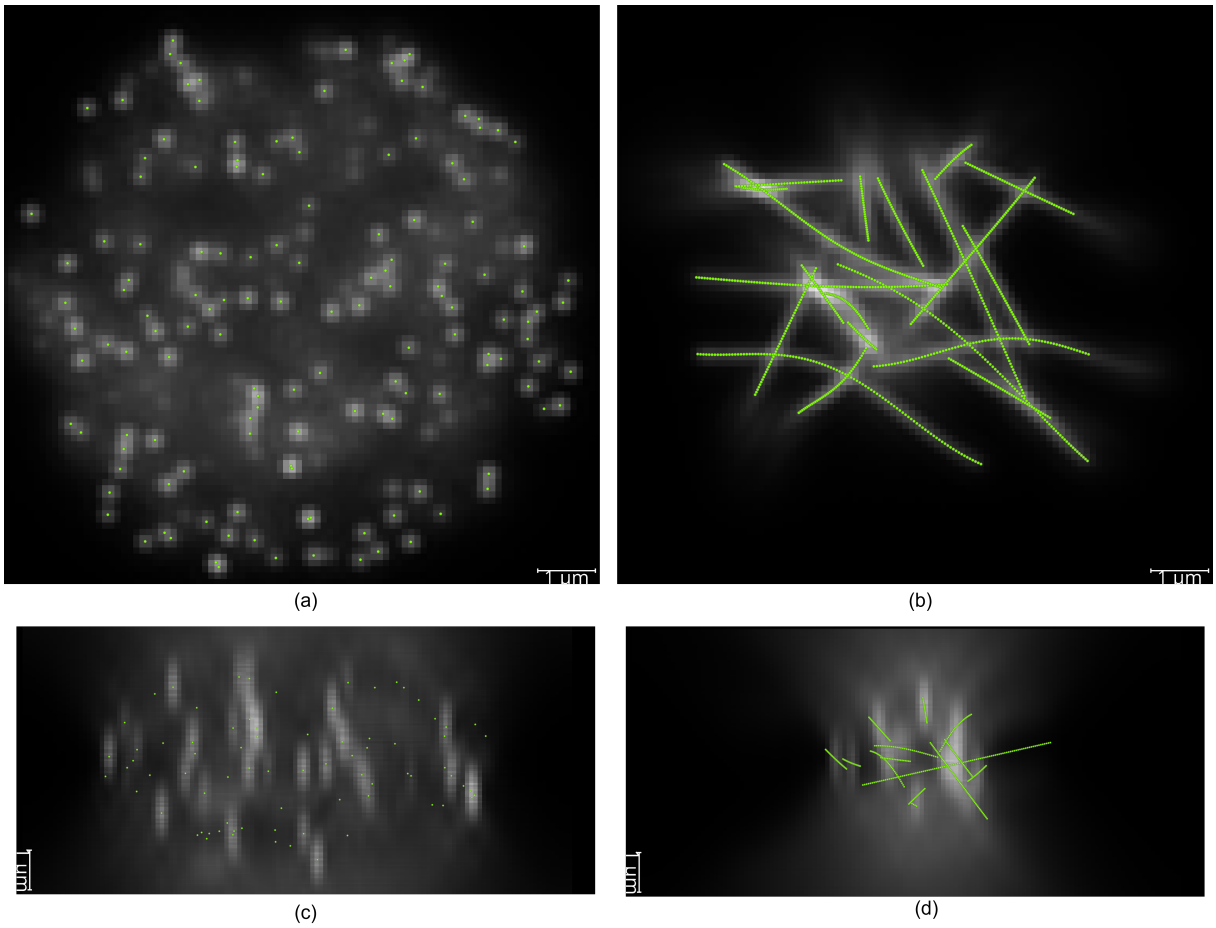


Figure 3-13. The superposition of a spheroidal sample point-cloud and the corresponding raw image in x-y (a) and x-z (c) plane. Filaments overlapped with the point-cloud in x-y (b) and x-z plane (d). Point-cloud samples are shown in green.

3.6.2. Speed of phase tilt calculation

Calculation of tilted phases is the overall rate limiting step in running napari-generic-SIMulator to simulate raw SIM image stacks. To mitigate this time-consuming part, various numerical modules are used in python to speed up calculation. Cupy is a GPU-accelerated version of numpy that processes data on a Cuda (NVIDIA, GeForce RTX 2080 SUPER) GPU (tested with cupy-cuda-11x). For devices which the GPU is not available on, the Torch package can be used, that provides many of the numpy capabilities and can often perform better on a CPU. In addition, the Torch also can accelerate its calculations on a Cuda GPU, if available (tested with torch v1.13.1+cu117). An example speed test is applied on a simulated spheroidal point-cloud with 500 points and 135 frames and a drift of 50nm in the z-direction for each frame. The

PSF is simulated for the scalar case. The time calculating the tilted phases with different pixel numbers of the simulated image stack is shown in Table 3-3. All time values in the table are the average of the second, third and fourth simulation because the preparation overhead, especially for GPU-simulations, is often seen in the first iteration. Clearly there is a demonstrable benefit of using a GPU for these simulations.

Pixel number / Time/s / Acceleration	128 x 128	256 x 256	512 x 512
Numpy	2.76	8.59	33.19
Torch_CPU	1.57	5.28	22.09
Torch_GPU	0.50	1.04	3.10
Cupy	0.47	0.46	0.75

Table 3-3. Tilted phase calculation time in second. This only illustrates a simulation example.

3.7. napari-generic-SIMulator Results

Some example PSFs and OTFs are presented in this section. In addition, stack of raw SIM images is generated and then proceed using napari-SIM-processor to obtain super-resolved images.

3.7.1. Simulated PSFs and OTFs

By utilising napari-generic-SIMulator, in different (scalar, vectorial rigid and vectorial flexible) scenarios, both PSFs and the OTFs of an imaging system can be generated as a 3-D stack. All parameters used in simulating scalar and flexible vectorial cases are shown in Figure 3-10, except for a large pixel number of 512 to prevent interference at the image edges. For the rigid simulation, the illumination and detection NA and the refractive index are all set as the same as 1.1 to create an extreme example: a light-sheet case. As shown in Figure 3-14, this scalar model explained in section 3.5.2 can be used in the napari-generic-SIMulator to generate a 3D scalar PSF.

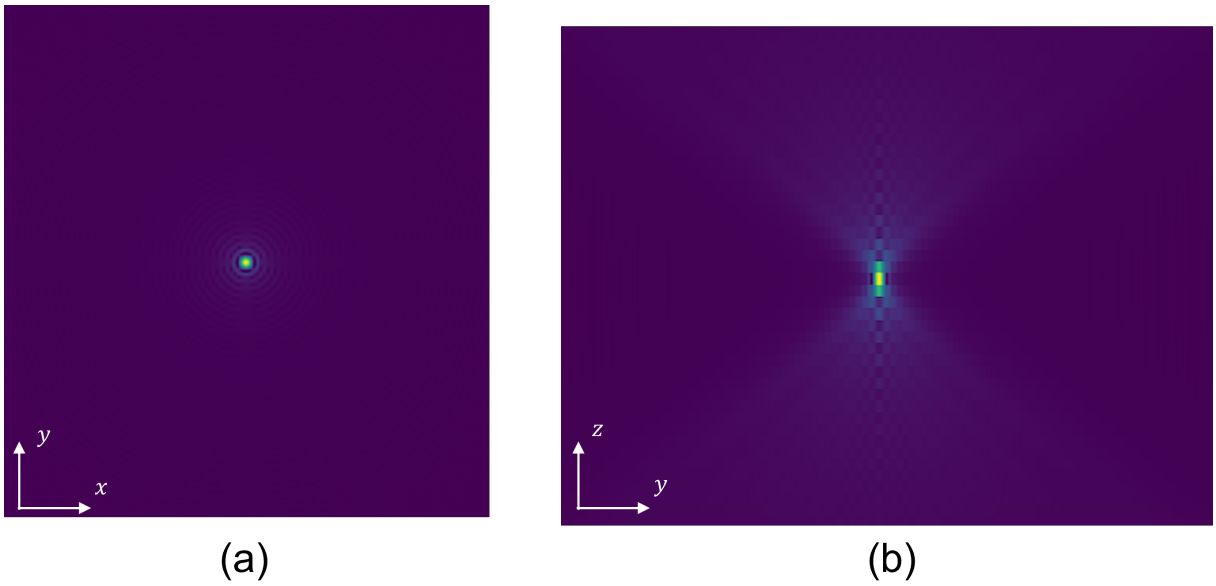


Figure 3-14. The slice of a scalar PSF in $x-y$ plane (a) and $y-z$ plane (b).

To study the PSF of a high NA system, the parameters here are chosen to simulate a light-sheet situation. The illumination and the detection NAs are set as the same as the refractive index of the medium, so the half angle is ninety degrees. With the same parameters apart from the 'psf' choice in napari-generic-SIMulator, vectorial PSFs can be simulated. Slices of the PSF in two planes are presented in Figure 3-15 (a) and (b).

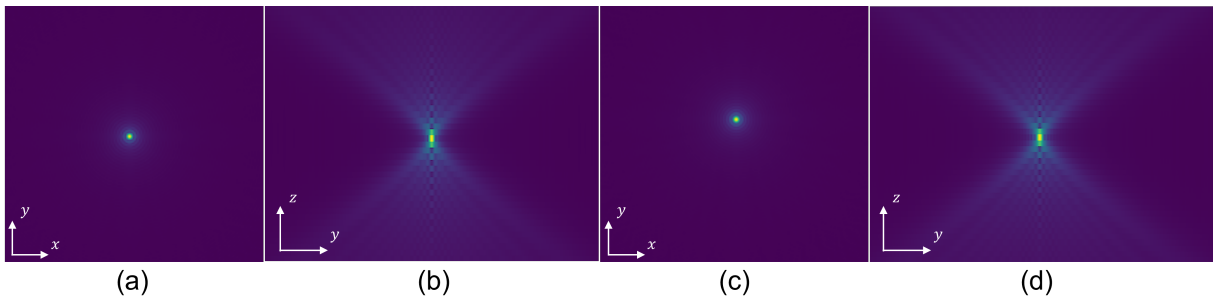


Figure 3-15. Vectorial PSFs: The PSF stack of flexible dipoles in $x-y$ (a) and $y-z$ plane (b). The PSF calculated with rigid dipoles in the horizontal (c) and vertical plane (d). Apart from the PSF calculation, all parameters are the same as those used to simulate the scalar PSF.

It is hard to visually observe differences between the vectorial PSFs in flexible and rigid scenarios (Figure 3-15) and the scalar PSF (see Figure 3-14). However, the corresponding OTFs demonstrate the differences in distributions of detected spatial frequencies more clearly, especially in the case of high NA system simulations.

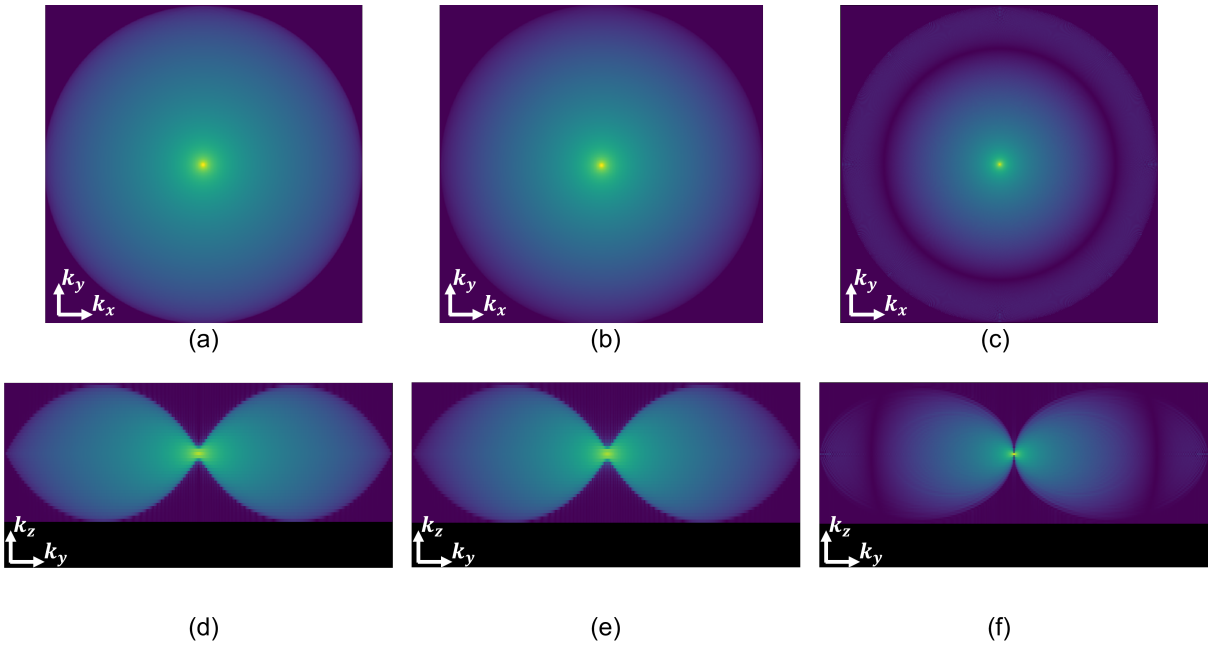
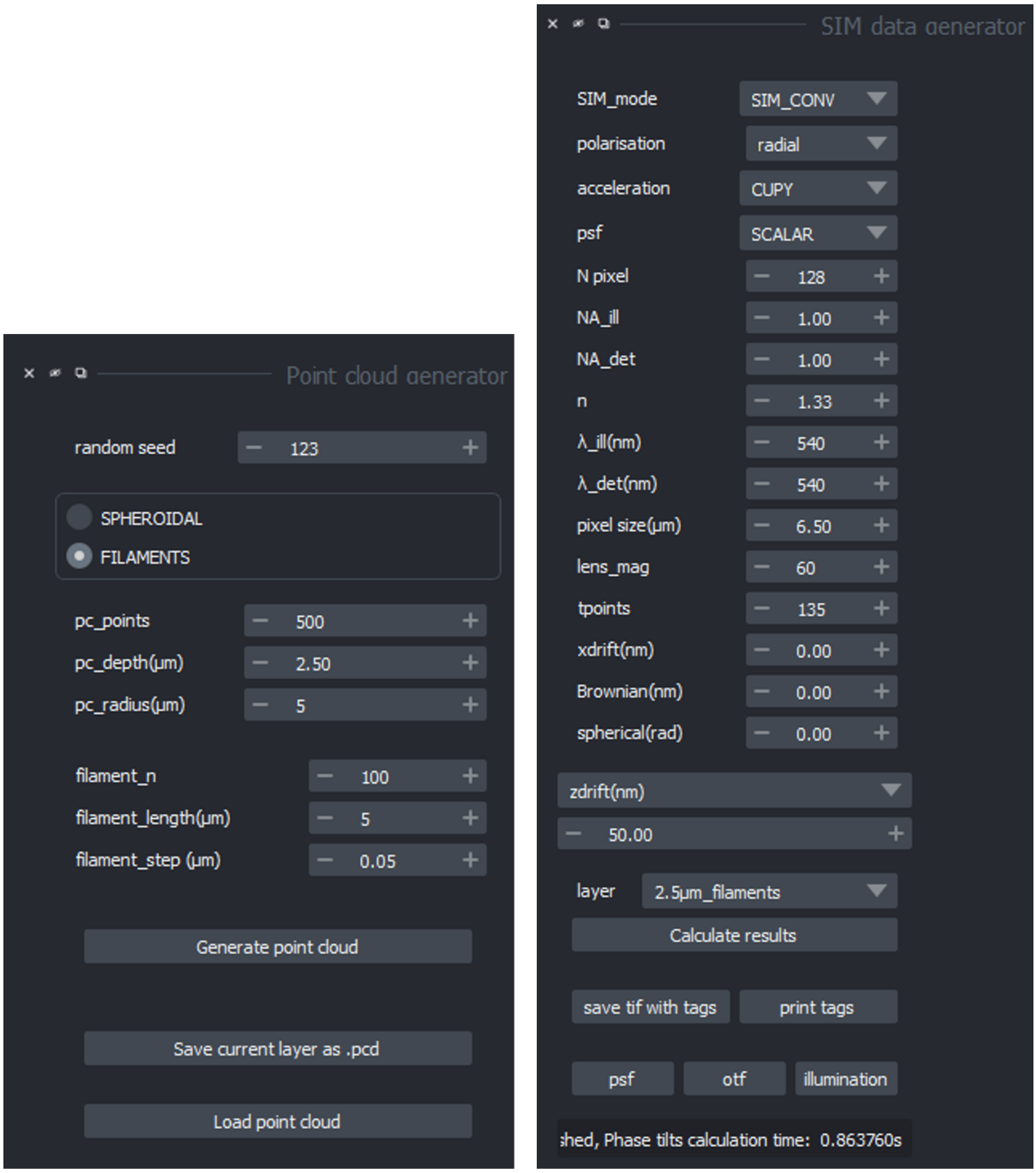


Figure 3-16. Scalar OTF in $k_x - k_y$ (a) and $k_y - k_z$ planes (d). The horizontal (b) and vertical (e) slice of the flexible OTF. There is an anulus of zero in the rigid OTF with z-polarised excitation that can be seen slices in the two planes presented in (c) and (f).

The corresponding 3-D OTF can be calculated and generated as 3-D models using napari-generic-SIMulator. Figure 3-16 presents k_x, k_y and k_y, k_z planes for each of the three 3-D OTF models, scalar ((a), (d)), vectorial with flexible fluorophores ((b), (e)) and vectorial with rigid fluorophores ((c), (f)). As shown in Figure 3-16 (c) and (f), there is a ring of zero in the simulated OTF when the polarisation of dipoles is excited along the optical axis, which means that the OTF calculated with rigid fluorophores is more complex than the other two situations.

3.7.2. Processing simulated raw SIM images

A raw image stack is generated using napari-generic-SIMulator, and the parameters are shown in Figure 3-17.



(a) (b)

Figure 3-17. Parameters of the simulated raw image stack in (a) point cloud generator and in (b) SIM data generator.

They can be aligned and reconstructed to show the accuracy of the simulated illumination generated by the SIMulator. A generic SIM processing platform, also another napari plugin: napari-SIM-processor [58] is written by Prof. A.Bassi and Prof. M.A.A.Neil for processing SIM raw image stacks. The stack of raw images is firstly reshaped according to the numbers of z, phase and angle, and is then processed in the Fourier space to find the frequency carriers of

the illumination patterns before the final reconstruction for super-resolution images. Figure 3-18 displays the GUI of the plugin, showing the reshaping and the reconstruction widgets.

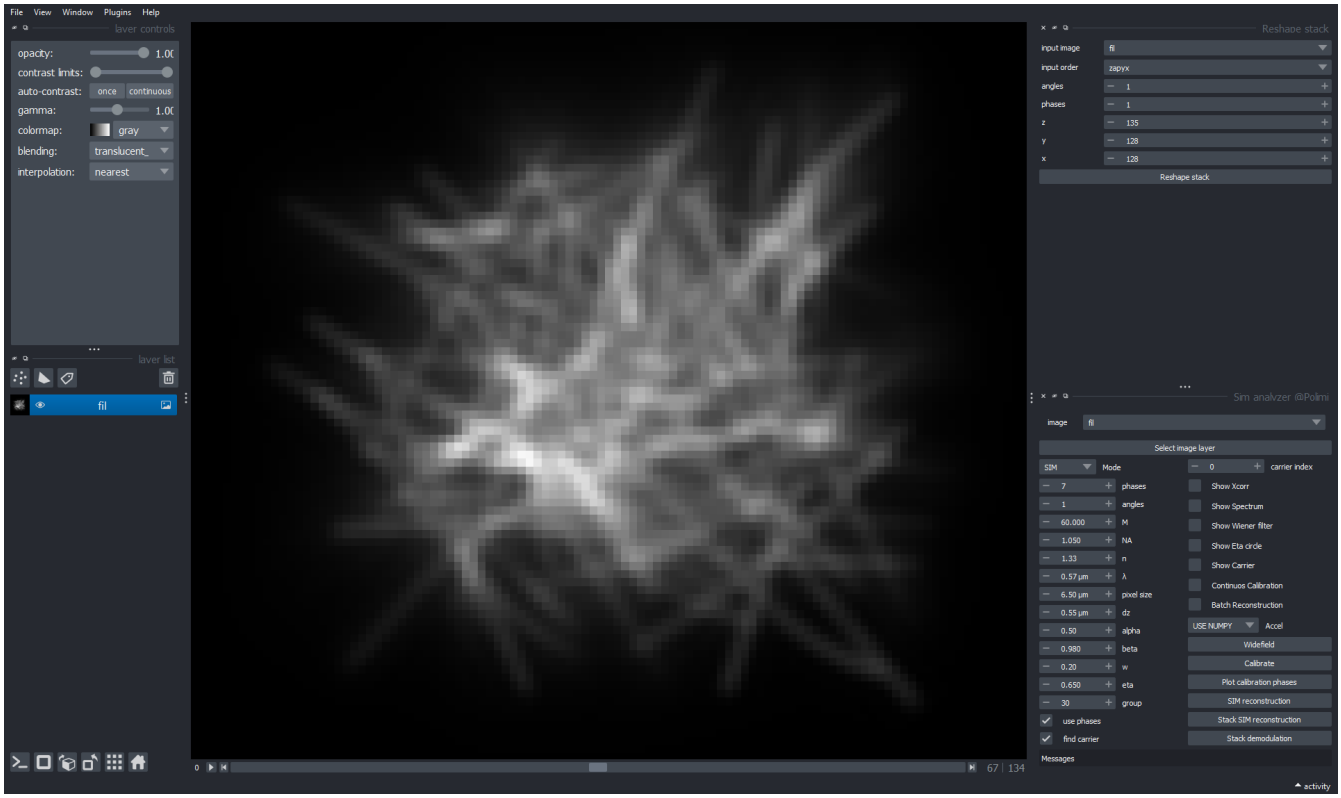


Figure 3-18. napari-SIM-processor GUI. The loaded image stack is the filament-point-cloud simulated above. Parameters shownin the figure are default values, which need to be changed according to the raw images.

The stack of simulated filaments mentioned above in Figure 3-17 is illuminated with conventional 2-D sinusoidal patterns and is processed by the napari-sim-processor as an example to show the performances of both the simulated data and the processor. The 135 raw frames in this stack are reshaped by ‘Reshape Stack’ widget into groups based on the number of phases, angles and z locations as 3, 3 and 15, respectively. Figure 3-19 illustrates images at different z positions for the first phase and the angle step, where the result of scanning the sample through focus along the optical (z) axis can be seen.

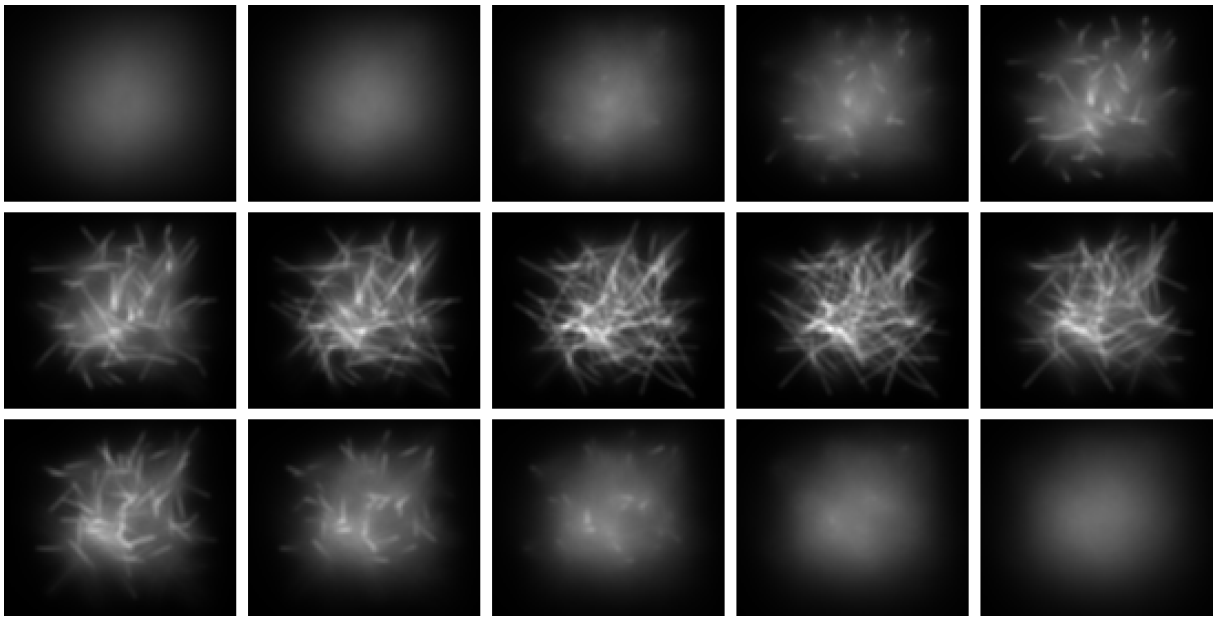


Figure 3-19. Raw images along z axis with 1st phase and 1st orientation. Top row: 1st -5th frame. Middle: 6th -10th frame. Bottom: 11th -15th frame. The distance between displayed frames is 50nm.

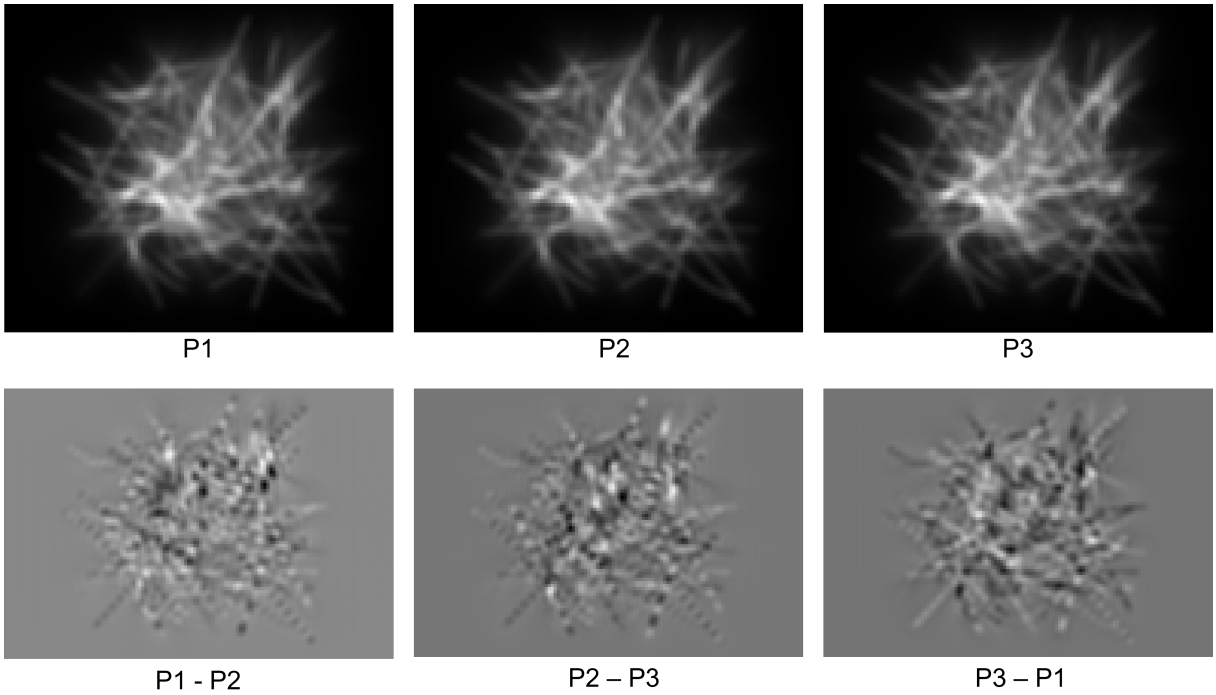


Figure 3-20. Raw frames with different phases and the differences. Top row: 3 raw images with 3 phases: P1, P2 and P3 at the first angle: A1. Bottom: differences between frames.

While keeping z at the middle position (the 8th frame), the sample illuminated with three phases and angles are displayed in Figure 3-20 and Figure 3-21, respectively. In the figures, in spite of the similar simulated camera images in top rows, the bottom difference images clearly

demonstrate the effects of the changing illumination. It is easy to observe the difference between frames by scrolling through the frame set along the phase or angle axis in the napari viewing window.

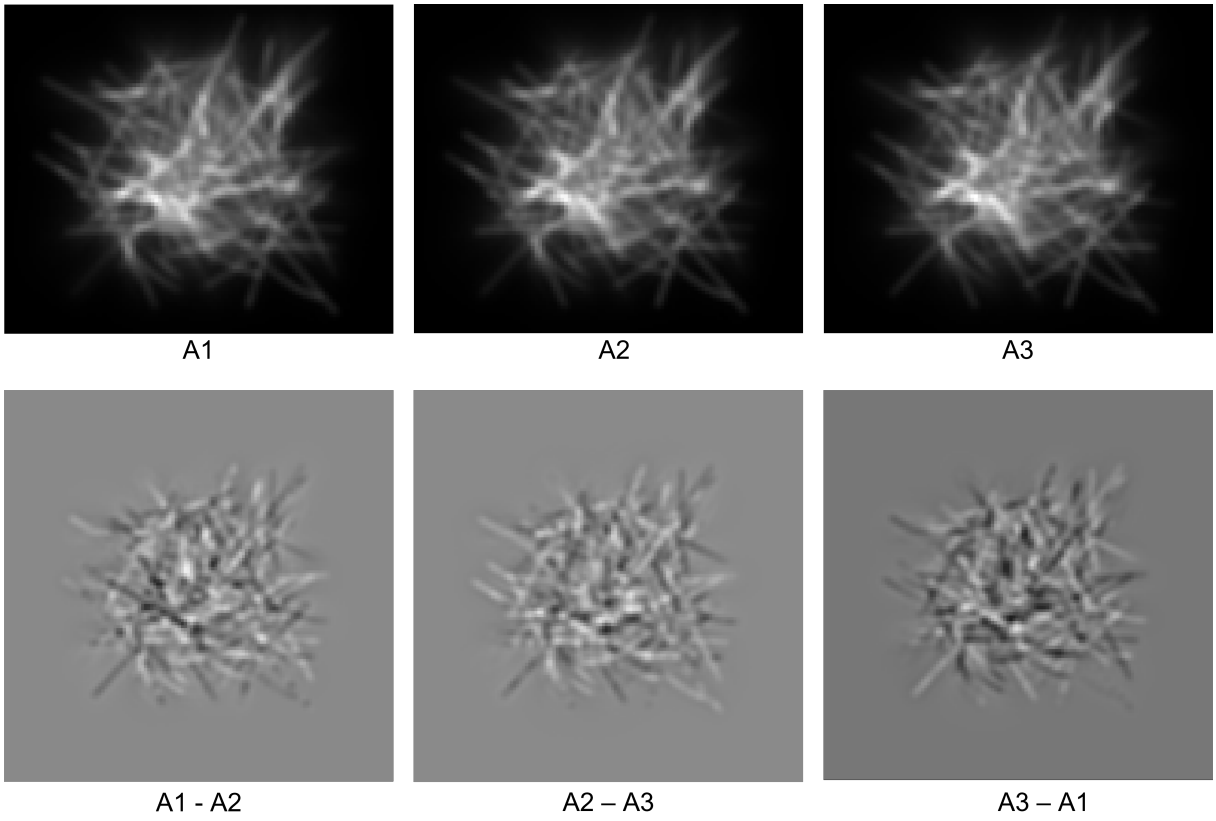


Figure 3-21. Three raw images at different angles (top) and the frame-differences (bottom) at the 2nd phase (P2).

Given the SIM raw stack with a correct shape, the frames can be calibrated to find the frequency components of the illumination patterns. In the plugin, the cross-correlation of a raw image is calculated to determine the locations of carrier frequency components from the structured illumination patterns. Since the frequencies are conjugate (symmetric in the plane) and there are three orientations of the illumination pattern for a 2-D conventional SIM case, three spots can be seen in the cross-correlation (as shown in Figure 3-22). In the figure the yellow circles annotate the frequencies at different angles, with the red ones showing the current checking. The two values displayed with circles highlighting the carrier positions are the amplitude and the phase of carriers calculated from the cross-correlation. The blue circle shows the boundary of the incoherent imaging region of support in frequency space, while the green circle shows the lower limit of the search the area used to find the range where we find the carrier orders and can be adjusted in the processor. The component at three angles in Figure 3-22 are all at the blue circle because the raw image is simulated and is able to double

the resolution in the processed image compared to that in the raw images according to SIM theory.

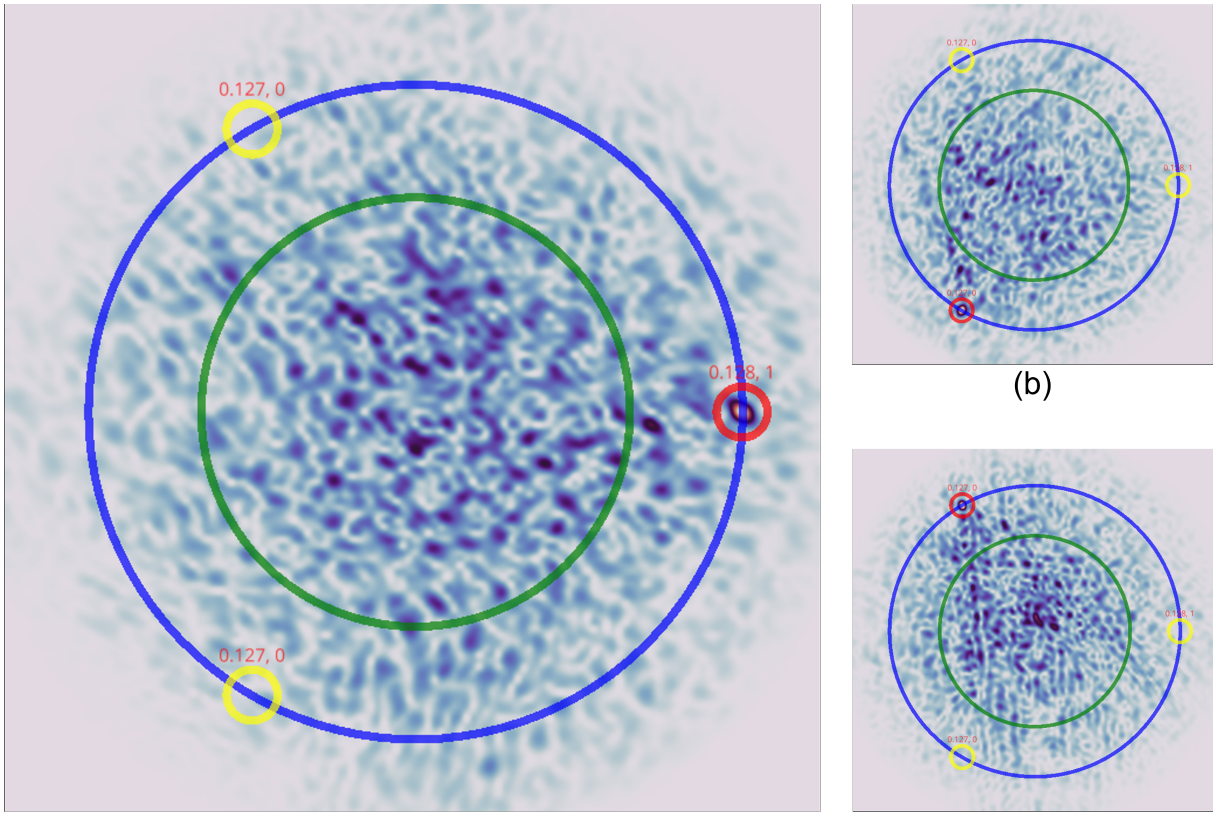
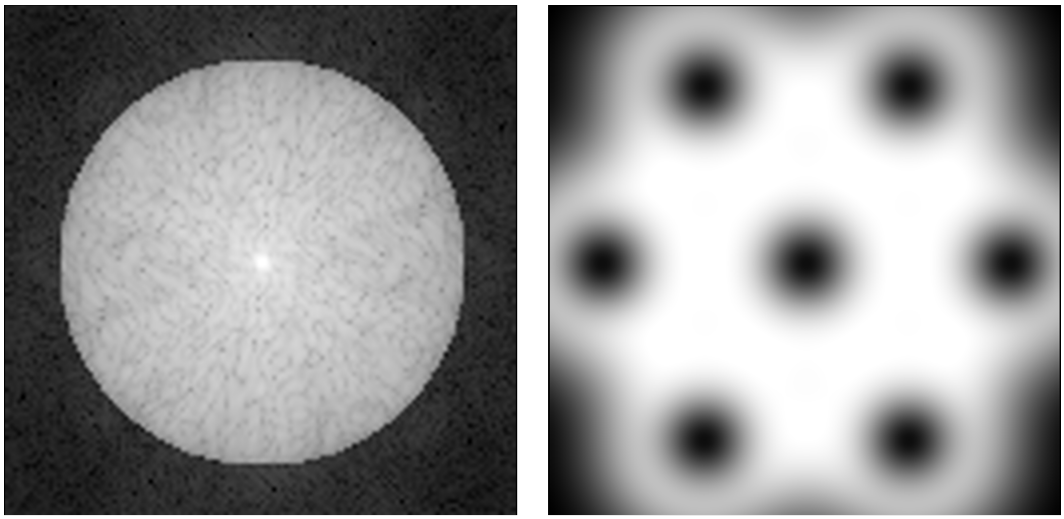


Figure 3-22. Cross-correlation of the raw image that is at the middle of the z-stack. The red circle in (a) (b) and (c) denotes the location of the frequency order at the 1st, 2nd and 3rd angle, respectively.

While processing raw images acquired from real experiments, it is generally not possible to reach the twice cut-off value due to noise and other practicalities in the imaging system. As shown in Figure 3-23, the spectrum of the raw images and the and the Wiener filter calculated for the reconstruction process can also be shown in the plugin. The region of the spectrum is the same as that limited by the blue circle in Figure 3-22, and the Wiener filter symmetric, and consistent with the simulated frequency components, illustrating that a correct calibration has been achieved.



(a) (b)

Figure 3-23. Two calibration results provided by the Napari-SIM-processor: the raw image spectrum (a) and the Wiener filter used in reconstruction (b) of the enhanced resolution image.

All these calibration diagnostics verify the correct phases and orientations of the simulated illumination, and also mean that super-resolved images of the simulated stack can be reconstructed in the plugin. It should be borne in mind that the raw frames include a drift in the z-direction of 50 nm per frame during the sequential acquisition to simulate the movement of the sample through focus. These can be processed either using a conventional SIM reconstruction algorithm or the batch method introduced in [38] that applies a central filter on the axial spectrum to remove artefacts due to the sample moving during acquisition.

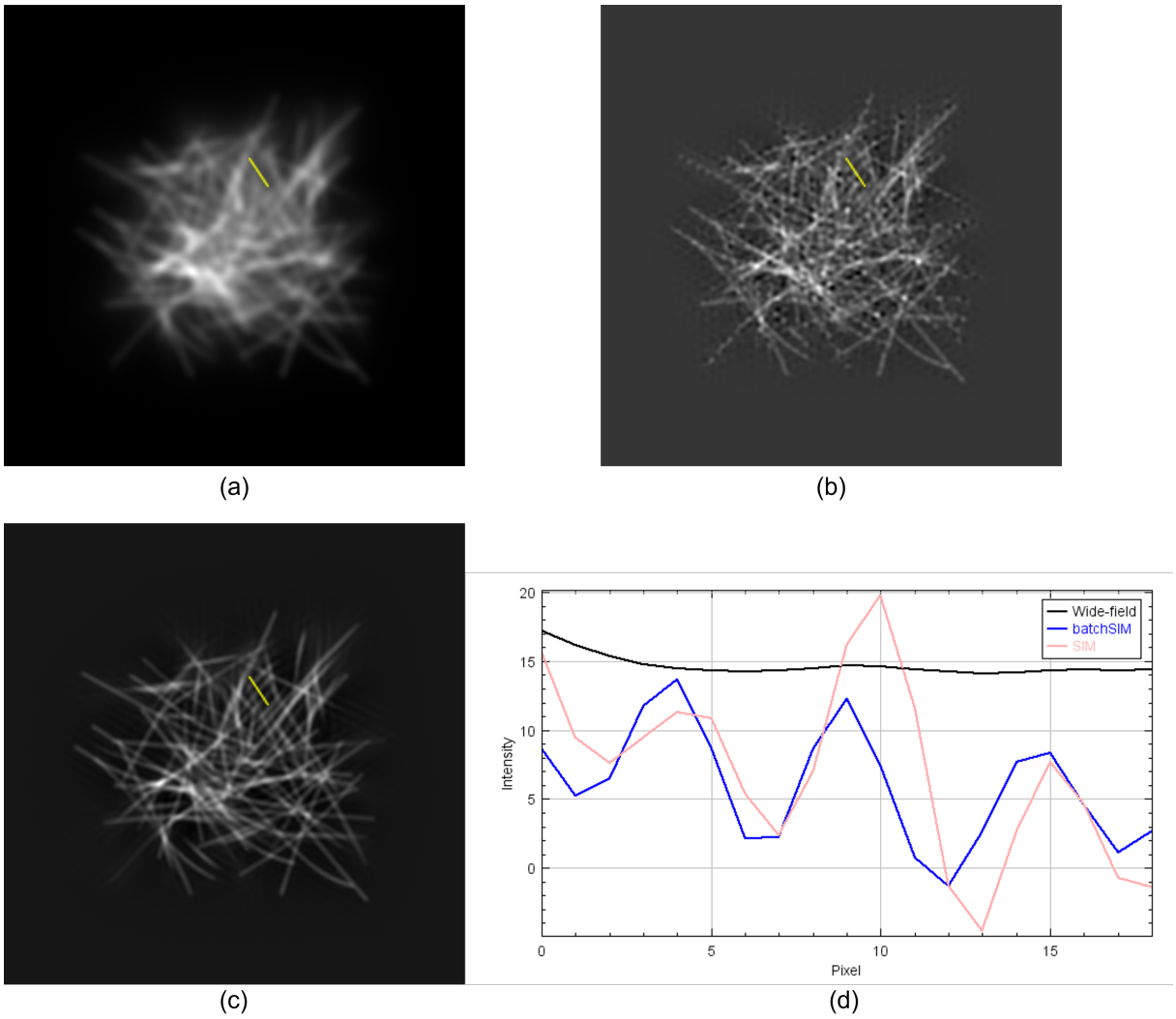


Figure 3-24. (a) The wide-field image. Reconstruction of the simulated raw image using conventional SIM (b) and batchSIM (c). (d) Intensity profile along yellow lines in (a), (b) and (c).

The reconstruction result of the eighth z-slice in the raw filaments stack mentioned above is demonstrated in Figure 3-24, compared to the wide-field image (a) which is the average of the corresponding nine frames, reconstructions using the standard SIM method (b) and the batchSIM (c) are highly resolved. This can also be seen in the intensity profile along the yellow lines (d) that the two SIM reconstructed results show three clear peaks while no peaks can be found in the wide-field one. The peaks using the batchSIM technique (the blue line) are clearer than those of the conventional SIM reconstruction (pink) in (d), and significant of artefacts can be seen in (b), that are caused by the continuous z-drift and are effectively alleviated by batchSIM as shown in (c).

3.7.3. Results from other SIM geometries

Figure 3-25 shows similar results that can be generated by the napari-generic-SIMulator.

All raw images are generated using napari-generic-SIMulator, with the parameters being the same as that in Figure 3-17, apart from SIM geometries and beam polarisation states. All reconstructions are realised using ‘Stack SIM reconstruction’ with ‘Batch Reconstruction’ of napari-sim-processor.

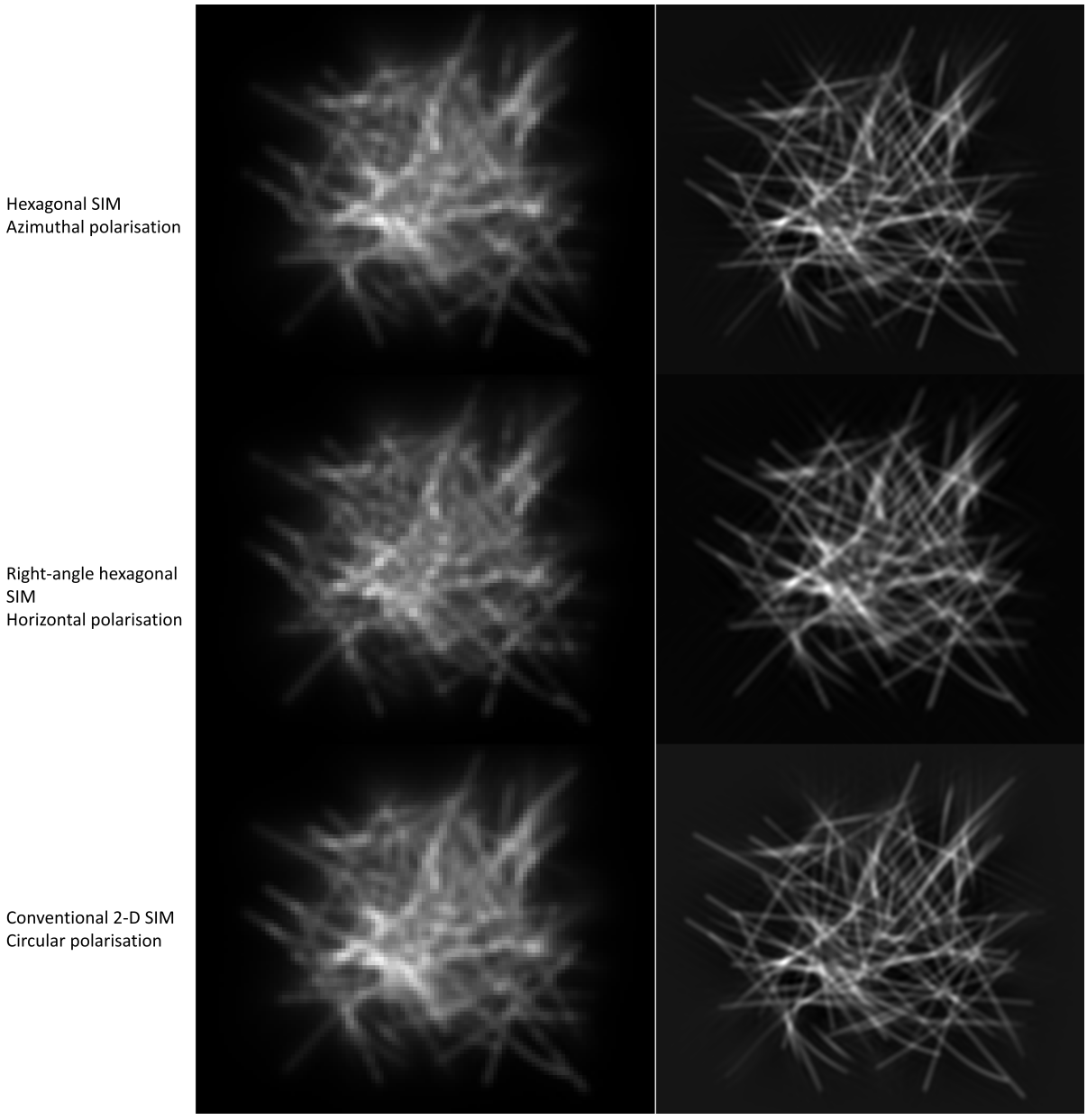


Figure 3-25. Raw images and reconstructed frames of different SIM geometries with various beam polarisation states. Left column: a raw image. Right column: super-resolved images.

The structure of the code is split into a base class that implements all the fundamental calculations required to simulate SIM images. Implementation classes for the different SIM geometries are then derived from this base class, where the construction of the specific illumination patterns is handled by the derived classes while the calculation of the resulting images is handled by the base class. It is by this mechanism that the plug-in becomes “generic” as it is easy to implement a new SIM mode, by writing a new derived class. By way of example, Figure 3-25 shows example of simulations and reconstructions for a range of different SIM geometries as described in Figure 3-2 and Figure 3-3.

3.8. Chapter summary

Based on the theory and computation of the simulation of SIM imaging, a bespoke napari plugin of the simulation of SIM imaging is introduced. The basic parameters such as the wavelength of the laser source, the magnification of the objective and the pixel size of the camera and the numerical aperture of both illumination and image detection can be changed to simulate different SIM systems. Samples used for SIM imaging are simulated as point clouds of a spheroidal distribution of points or filaments with adjustable parameters for various specimens. Structured illumination patterns are calculated based on amplitudes and phases of Jones vectors, at each spot in the point-cloud sample. The structured illumination can be chosen from conventional two-beam, hexagonal, right-angle hexagonal and conventional 3-D modes, with different polarisations of beams including horizontal, vertical, azimuthal, radial, and circular states. Both scalar and vectorial PSFs and OTFs can be simulated. Raw SIM stacks simulated by napari-generic-SIMulator are presented and one of them is aligned and analysed with napari-SIM-processor. The calibration and reconstruction results demonstrate the accuracy and the similarity to a real SIM stack of the raw frames and test the napari-SIM-processor’s capability of mitigating artefacts introduced by drifts along z direction.

Apart from the exhibited examples, napari-generic-SIMulator can choose different basic parameters to simulate real SIM scenarios, tailor different point-clouds, generate user-definable aberration in detection path such as the spherical aberration, add artificial phase and angle errors in the illumination and also introduce drifts in three dimensions to samples. With different choices of SIM modes, arbitrary SIM illumination geometries can be produced as well. Combining with useful functions of viewing window in napari, illumination pattern stacks, three dimensional PSFs and OTFs can be displayed in the plugin. To be more realistic, some readout and shot noise can be added to present real SIM raw images.

We foresee that this napari plugin will be a capable tool to generate raw data for testing SIM reconstruction algorithms and can efficiently reveal the artefacts in SIM reconstruction caused by sample movement including drifts and Brownian motions. The clear and specific structure of the Python code of the plugin also guarantees the possibility to simulate other versatile types of samples and SIM illumination geometries in the future of developing SIM. In addition, a big number of raw stacks can be generated on CPU or GPU at a high speed to provide enough models for developing SIM with machine learning.

3.9. Appendix: Nyquist sampling in the pupil plane and the PSF plane

The main part of simulating an imaging system is to calculate the PSF according to the parameter simulation. To this end, the correct spacings in lateral planes need to be determined. Depending on the camera's pixel size and the objective's magnification in the system, the lateral spacing dxy in the detection plane (also the camera plane) can initially be calculated as:

$$dxy = \frac{pixel\ size}{magnification} \tag{3-47}$$

, assuming the spacing in x and y directions is the same. However, in order to avoid aliasing, the sampling rate should satisfy Nyquist sampling theorem. Therefore, the maximum of Nyquist sampling dxy_n at the sample is calculated as half the shortest detection distance d_{det} in the system:

$$(dxy_n)_{max} = \frac{1}{2} d_{det} = \frac{1}{2}\left(\frac{\lambda_{det}}{2 \times NA_{det}}\right) \tag{3-48}$$

, where NA_{det} and λ_{det} is the detection numerical aperture and the detective wavelength, respectively.

In our simulation with finite numbers of spacings along x (or y) directions,

$$N * dxy = N_n \times (dxy_n)_{max} \tag{3-49}$$

, where N is the number of x (or y) values that is adjustable by us for different simulations, and N_n is the corresponding number of pixels at the Nyquist sampling.

The N_n we can take is thus expressed as $N * dxy$ / $(dxy_n)_{max}$, which is taken as the minimum even number by:

$$ceil\left[\frac{N_n}{2}\right] \times 2 \quad \text{3-50}$$

in the code, where $ceil[x]$ maps x to the least integer not greater than x. The xy-spacing can be corrected to:

$$dxy_n = \frac{dxy \times N}{N_n} \quad \text{3-51}$$

. In this way, the lateral spacing at the sample can assure Nyquist sampling and to be large enough to avoid oversampling.

The other important lateral plane is the pupil plane, and it should be emphasised that the pupil is not the entrance or exit pupil in the system but represents a 2-D Fourier transform of a field. In terms of the spacing in the pupil plane, given that the wave number of the detection light k_0 is defined as $\frac{2\pi n_r}{\lambda_{det}}$, where n_r is the refractive index, the maximum detectable lateral frequency $k_{r_det\ max}$ can be calculated as

$$k_{r_det_{max}} = k_0 \times \frac{NA_{det}}{n_r} \quad \text{3-52}$$

, which defines the maximum range of coordinates in the pupil plane. The factor by which pupil plane oversamples the coherent PSF data can also be calculated as the factor between the spacings of the detection and the sample plane:

$$\frac{k_{pupil}}{k_{r_det_{max}}} = \frac{d_{det}}{dx_{yn}} \quad \text{3-53}$$

. Because the detected emitted fluorescence is an incoherent process, the spacing dk in the pupil plane can then be expressed as

$$dk = \frac{k_{pupil} \times 2}{N_n} \quad \text{3-54}$$

. From 3-53 and 3-54, dk can be calculated as:

$$dk = \frac{2 \times d_{det} \times k_{r_det_{max}}}{dx_{yn} \times N_n} \quad 3\text{-}55$$

. The coordinates k_x and k_y values are determined as from $-dk * \left(\frac{N_n}{2}\right)$ to $dk \times \left(\frac{N_n}{2}\right) - dk$, keeping the interval as dk (they are defined as a meshgrid in the Python code). To make it easier to describe the pupil, the radius k_r is written as:

$$k_r = \sqrt{{k_x}^2 + {k_y}^2} \quad 3\text{-}56$$

. Similarly, in z direction, the maximal Nyquist sampling is calculated:

$$(dz_n)_{max} = \frac{1}{2} \times \left[\frac{\lambda_{det}}{n * (1 - cos\theta)}\right] \quad 3\text{-}57$$

, where θ is the half angle of the acceptance cone of the objective as $\theta = \arcsin\left(\frac{NA_{det}}{n}\right)$. According to the spacing, the number of pixels in z direction Nz_n is taken as

$$Nz_n = ceil\left[\frac{z_{range}}{(dz_n)_{max}}\right] \times 2 \quad 3\text{-}58$$

, where z_{range} is determined according to the sample. k_z coordinates can be set as an array from $-\frac{N_{z_n} * dk_z}{2}$ to $\frac{N_{z_n} * dk_z}{2}$ with a dk_z between values, where dk_z is $\frac{\pi}{z_{range}}$. The dz_n is then taken as

$$dz_n = 2 \times \frac{z_{range}}{Nz_n} \quad 3\text{-}59$$

Chapter 4. Correction for the flatness of R11-2K SLM

This chapter illustrates how a practical issue of flatness in the spatial light modulator (SLM) results in distorted diffraction orders and a method for solving this by determining the flatness as an aberration and correcting it in the hologram design. In a SIM system, an SLM is used to diffract the laser light into a number of beams that interfere to generate structed illumination patterns. In order to block unwanted diffracted orders a pinhole mask is used and accordingly, the SLM's performance is important since the quality of the required diffracted orders must be good (small) enough to pass through the pinholes and so this determines the quality of illumination patterns projected on the sample, and thus of the resultant super-resolved images. However, in practice, the surface of the SLM is found to not have the required flatness, and when the required diffracted orders are partially blocked by the pinholes this issue leads to an incomplete field of view (FOV) and artefacts in the reconstruction images. In this chapter, the problem is demonstrated, and an approach is introduced to measure and correct the aberration caused by the non-flatness of the SLM. Results of both simulations and experiments are presented.

4.1. The distorted diffraction order

Ferroelectric liquid crystal on silicon (FLCOS) SLMs can be used to generate diffracted beams by displaying holograms. For a FLCOS SLM, an applied electric field drives the ferroelectric liquid crystals to change the polarisation of the incident light.Initial work on the system used an SLM (R3-QXGA, Forth Dimension Displays) which was connected to a computer via a video card [38]. In this work the currently used model (R11-2K, Forth Dimension Displays) however, is connected via USB. The USB driver offers more flexible triggering and display options than the video card driver, which makes on-the-fly phase recovery accessible and more efficient. Additionally, there are more pixels (2048 x 2048) in the R11-2K model than the QXGA (1280 x 1024) SLM. It also has a smaller pixel size, providing for a more refined hologram, and better control of artefacts in reconstructed images (as described in chapter 6).

As presented in Figure 2-1, a relatively simple set-up is implemented to measure the intensities of diffraction orders generated by the SLM. A laser with a low power value is expanded and collimated onto the face of the SLM, displaying designed holograms loaded onto the SLM system. The light is diffracted into a few orders that are focused into the imaging plane of the camera after passing through a neutral density (ND) filter. The holograms are produced by the

DBBI method explained in Chapter 5, with the number of beams being only one since the measurement on one diffraction order is sufficient to study the phase of a hologram. Still, three main orders can be imaged by the camera because there is a zeroth order beam locating at the centre and the conjugate of the first order symmetric opposite the desired +1 order position. However, the region of interest (ROI) of the camera is adjusted to only show one of the first diffracted beam (+1 or -1) at its centre. If the SLM was perfectly flat and the system had no aberrations then we would expect each of the diffracted orders to be a small spot, diffraction limited by the aperture of the SLM.

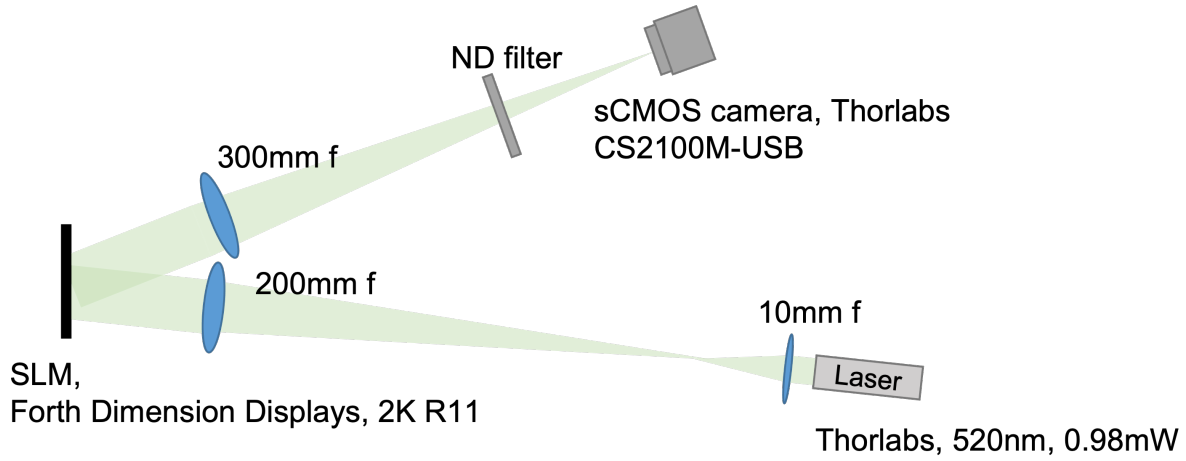


Figure 4-1. Optical layout for measurement of diffraction orders. A low-power laser is used as the light source.

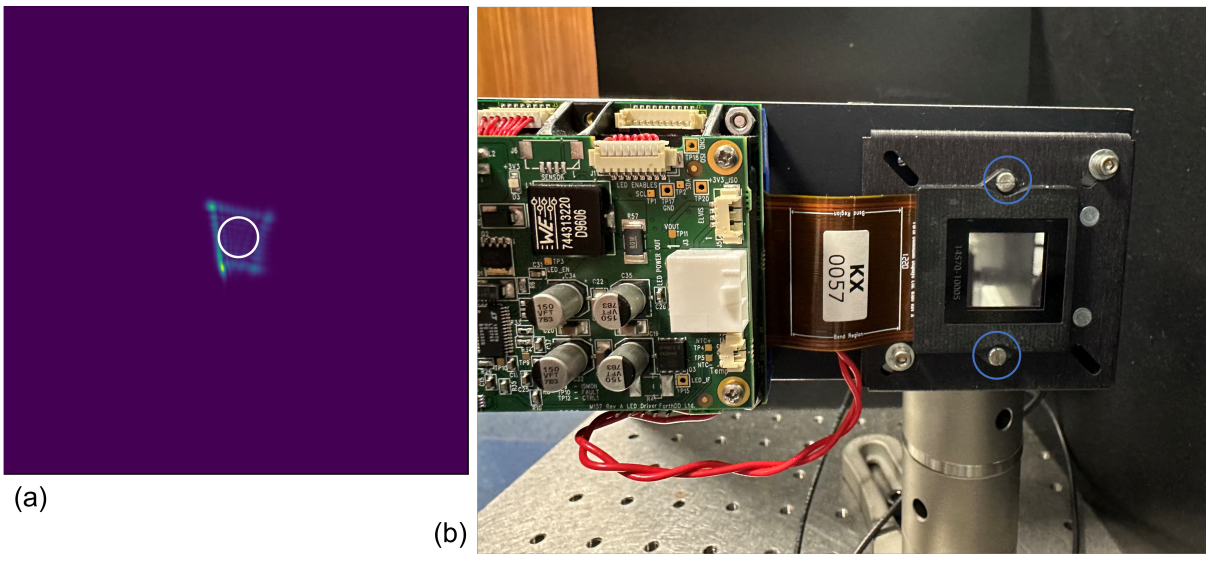


Figure 4-2. (a) The image of a diffraction order. The pinhole is denoted as a white ring. (b) The installation of the SLM. Two screws on the mounting frame are emphasised using blue circles.

Nevertheless, as shown in Figure 4-2 (a), the diffracted order is extended from a diffraction limited point to a roughly rectangular shape, demonstrating optical aberrations in the system. In practice, the shape of the diffraction order is mainly affected by the flatness of the SLM, as displayed in (b), which includes two screws on the mounting frame of the SLM that hold it against its temperature-controlled backplane. Specifically, the more these screws are tightened, the more seriously the beam is distorted. Typically, a pinhole mask would be used to restrict the illumination to just the required diffracted orders and to block other higher diffracted orders. The pinhole in our system for picking up desired diffracted orders has the radius as 100 μm, and is denoted as the white circle in (a). Obviously, the aberration introduced by the SLM is significant, the extended diffraction orders is largely blocked by the pinholes, as demonstrated in (a), making the field of view of incomplete (see results in Chapter 6). In practice, this issue cannot be entirely addressed by loosening the two screws since the liquid crystal screen of the SLM exhibits some inherent non-flatness. Additionally, it needs to remain securely fixed in position to ensure proper temperature control.

4.2. Phase recovery for measuring and correcting SLM aberrations

The plane where the SLM sits can be considered as the pupil plane of the system since the screen of the SLM restricts the laser to a square aperture, and the camera is located at the corresponding PSF plane. To compensate the inevitable aberration introduced by the curved SLM, the phase at the pupil plane needs to be characterised so that the displayed hologram and hence the electric field at the PSF plane can be controlled.

Based on Zernike aberration modes and measurements by a deformable membrane mirror, Booth et. al corrected the aberration in an adaptive confocal microscope system [66]. The wavefront sensor-less method is used to correct aberration in different microscope systems: multiphoton microscopes [67] [68] [69] and a multi-harmonic microscope [70]. In 2011 D. Debarre *et.al* demonstrated the estimation of the electric field in a system with a deformable mirror by applying the Gerchberg-Saxton (G-S) algorithm to measured intensity PSFs at three different focal positions [71], and they used this to characterise the influence function of the deformable mirror. This type of sensor-less method is significant in the development of adaptive optics [72] [73], and is used in other microscopy techniques such as single-molecule switching microscopy [74].

Similarly, based on PSF intensity measurements, we integrate an in-silico G-S algorithm with hardware feedback fitted to Chebyshev polynomials to estimate the aberration caused by the SLM and to recover its pupil phase in-situ.

4.2.1. Gerchberg -Saxton algorithm based on PSF measurements

The Gerchberg-Saxton (G-S) algorithm can be roughly explained as iteratively retrieving a phase by performing forward and inverse Fourier transforms between the image and its corresponding Fourier plane, respectively. During each iteration, constraints are applied, such as retaining the phase information and substituting the amplitudes with measured values. As displayed in Figure 4-3 (ignoring red annotations for now), an estimated pupil amplitude and phase is fast Fourier transformed to calculate an estimated PSF. The PSF amplitude is then substituted with a measurement while keeping the PSF phase. The substituted PSF is then inverse fast Fourier transformed to recalculate the pupil field. The phase of this calculated field is kept, while the amplitude is smoothed and limited to the pupil aperture. Practically, a smooth function is used to fit the pupil amplitude instead a replacement by measured values as this is expected to be an appropriate estimate given that it is simply an expanded laser beam as shown in Figure 2-1. This process is repeated until a standard is satisfied, which is set as 75 iterations in our algorithm.

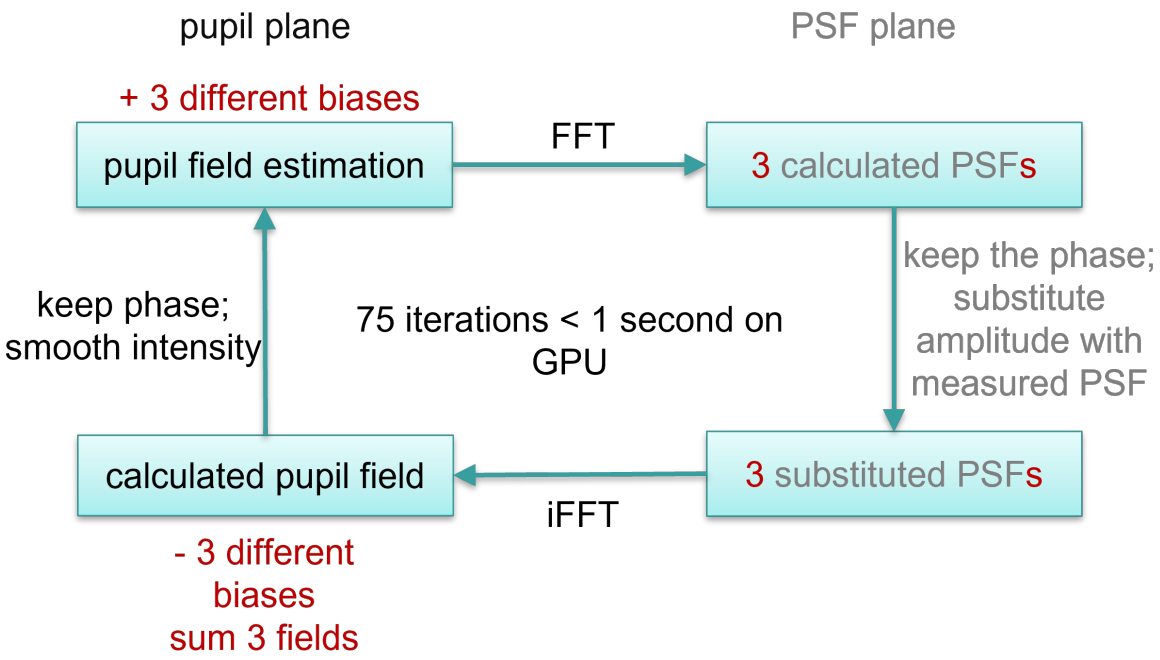


Figure 4-3. Gerchberg-Saxton algorithm. Red annotations are related to three introduced wavefront biases.

One problem with a general G-S phase reconstruction technique is that, for instance, the intensity PSF for positive defocus is the same as that for negative defocus, as too are those for positive and negative astigmatism. However, that for positive defocus plus positive astigmatism is different to that for negative defocus plus positive astigmatism. In the G-S algorithm in [71], which originated from the approach in [75], additional phase biases are added while measuring a set of PSFs [76], in order to ensure the unique estimation of pupil phase. In Debarre's case [71] these were simple defocus biases applied with a lens and piezo-mirror, while in our method we introduce wavefront biases holographically via the SLM pattern. To retrieve the pupil phase with the biases, as shown in the red notes in Figure 4-3, some extra steps are applied compared to a general G-S approach. In one iteration, with three wavefront biases added in the pupil field estimation, three PSF measurements corresponding to the same pupil are hence acquired, giving rise to three calculated biased fields by three inverse Fourier transforms. The number of added biases is empirically chosen as three , taking into account both the time required for algorithm calculations and PSF measurements. In the pupil plane, three biased phases are subtracted correspondingly, and all fields are then summed up to one field. Additionally, with different introduced aberrations, PSFs can be spread over more pixels rather than merely locating at one small area, which risks overexposure.

4.2.2. Problems with phase singularities in G-S results

The entire scenario is simulated with Python code to retrieve the phase of the pupil (SLM) plane that causes distorted PSF. To simulate both the aberration in the phase and the corresponding initial guess, Chebyshev polynomials (an orthogonal set over a defined -1 to +1 interval and so in a 2-dimensional form) are used to describe smooth functions covering a square pupil, similar to how Zernike polynomials are suited for a circular pupil. In addition, Lukosz modes can be used to represent aberrations [77] when the pupil is circular, especially when imaging low spatial frequencies [78].

Chebyshev polynomials can be used to provide a fit to an arbitrary function over a square aperture by two methods: by summing over weighted or sampled polynomials as explained later in section 4.3. As presented in Figure 4-4 (a), the continuous pupil phase φ of a non-flat SLM is simulated as a random combination of weighted Chebyshev polynomials $B_{w_{ij}}$ multiplied by the two-dimensional aperture mask M:

$$\varphi = \left(-7.5B_{w_{21}} - 12.5B_{w_{11}} + 17.5B_{w_{02}} + 27.5B_{w_{13}}\right) \times M \qquad 4\text{-}1$$

, where $B_{w_{ij}}$ (a Chebyshev polynomial term) is shown later in Figure 4-14. The Chebyshev polynomials are calculated across all pixels in the simulated pupil.

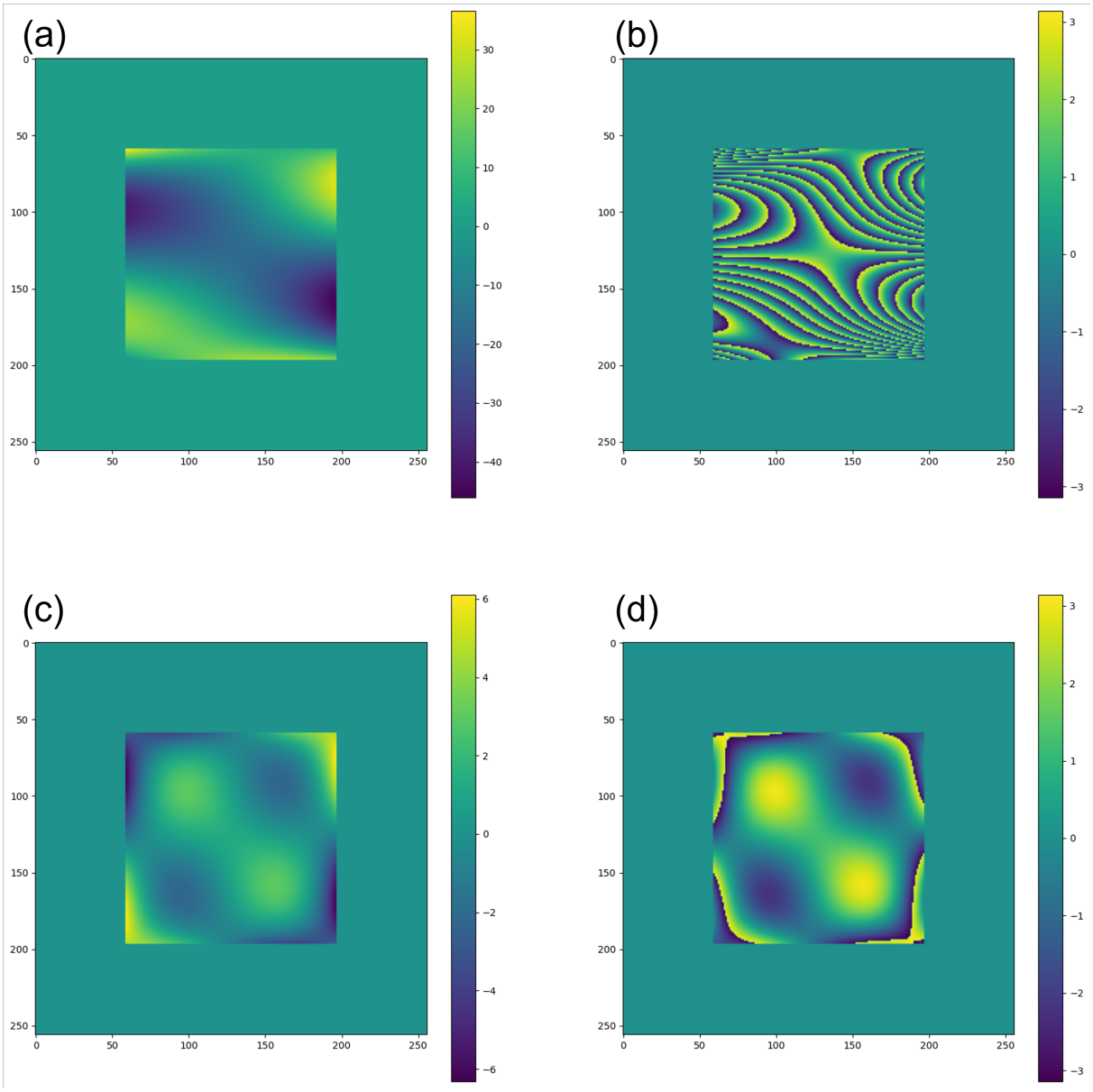


Figure 4-4. Preparation of the phase retrieval with Chebyshev polynomials. (a) Target pupil phase φ. (b) Wrapped pupil phase Φ. (c) Initial guess of the unwrapped pupil phase φ_g. (d) Wrapped estimated phase Φ_g. The pixel number of each phase is $N(256) \times N(256)$. The unit of the colour bars is radians.

The magnitude of the phase φ is outside of a normal range $[-\pi, \pi]$ radians since the sum of the random Chebyshev polynomials is arbitrary. To illustrate the change in φ more obviously, the phase is wrapped to be restricted to the range $[-\pi, \pi]$ radians as Φ,

$$\Phi = angle(e^{i\times\varphi}) \quad 4\text{-}2$$

, where the angle function $angle(x+iy)$ measures the angle between positive real axis and the line from the origin to the point (x, y) in the complex plane. As show in Figure 4-4 (b), there are curves in the wrapped phase Φ that contour 2π jumps and present the phase difference obviously. Similarly, as presented in Figure 4-4 (c), an initial guess of the pupil phase φ_g is generated consisting of another random mixture of $B_{w_{ij}}$.

$$\varphi_g = \left(-4.5B_{w_{11}} + 1.25B_{w_{22}} - 1.75B_{w_{31}} + 2.75B_{w_{33}}\right) \times M \quad 4\text{-}3$$

The wrapped estimated phase Φ_{g} can also be yielded according to the equation 4-2, as illustrated in Figure 4-4 (d). It is apparent that (d) is different from (b), showing that the target phase is unknown to the phase retrieval process. The amplitude of the pupil is simply defined as unity for simplicity as this the guessed value can also be random (but smooth). According to colour bars in Figure 4-4, the values of wrapped phases in (b) and (d) both lie in the range from $-\pi$ to π radians, while the range of unwrapped ones in (a) and (c) the range is approximately $[-45, 35]$ and $[-7, 6]$ radians, respectively.

Next, it's important to define the biased wavefronts introduced to the pupil plane to achieve the G-S algorithm mentioned in 4.2.1. Unlike classical names of Zernike polynomials such as defocus and coma, the indices of two-dimensional Chebyshev polynomials do not all correspond to terms for aberrations described across a circular aperture. However, $5 \times (B_{w_{02}} + B_{w_{20}})$ has a form corresponding to $(x^2 + y^2)$ and can be described as a defocus optical aberration, while $5 \times B_{w_{22}}$ has a form corresponding to $(2x^2 - 1)(2y^2 - 1)$ has characteristics similar to a higher order astigmatism and is referred to here as the astigmatism bias. In fact, the defocus with respect to the focal plane required to be considered [79] to avoid a dramatic drop in the signal-to-noise-ratio and an exceedingly long acquisition time for the measurements. As illustrated in the top row in Figure 4-5, the three aberrations in the phase are a positive defocus b_0, an astigmatism bias b_1, and a negative defocus b_2.

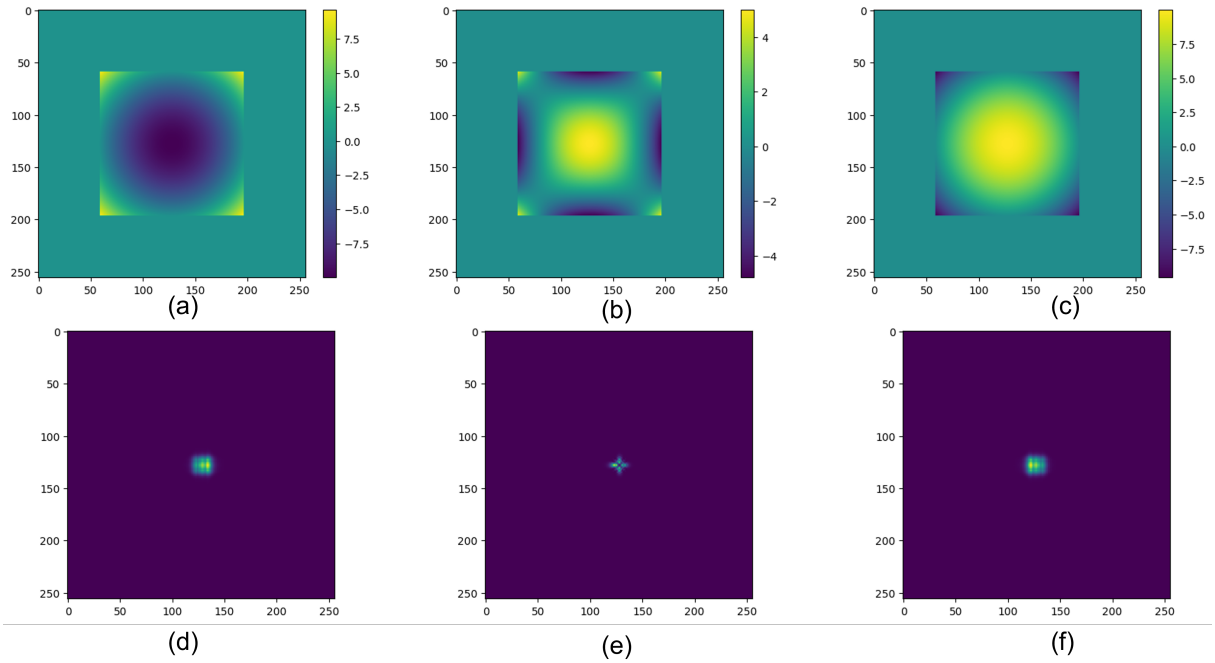


Figure 4-5. Three wavefront biases and corresponding PSFs. (a) The positive defocus b_0. (b) The positive astigmatism b_1. (c) The negative defocus b_2. These are all central symetric. (d) – (f): Corresponding PSFs.

With the biases and the distorted phase caused by the SLM, the PSF measurements can be simulated. The generation of holograms with the biased pupil field and loading the holograms to the SLM are ignored since the PSF at the imaging plane can be regarded as the Fourier transform of the pupil field. The real amplitude A of the simulated field in the pupil plane is then defined as

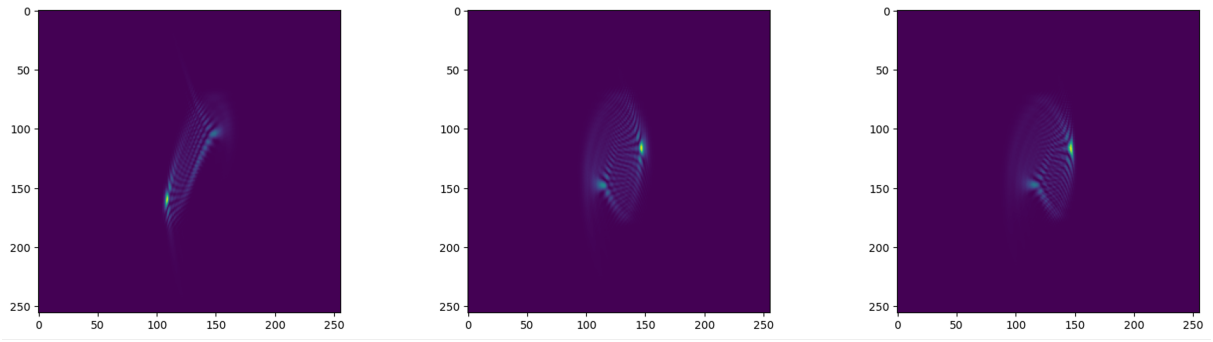
$$A = M \times e^{\frac{-(x-0.25)^2-2y^2}{2\times 1^2}} \qquad 4\text{-}4$$

to simulate an offset laser beam scenario by multiplying a shifted Gaussian function with the square mask. The PSFs corresponding to biases can be calculated by the Fourier transform, which are showed in the bottom row in Figure 4-5. Three aberration biases were chosen in order to spread the intensity in the camera plane, otherwise a diffraction limited unbiased and corrected PSF would risk saturating the camera. With three approximately equal intensity PSFs from the three biases a fuller dynamic range of the camera can be used for PSF measurements.

Each of three aberrations is added into the pupil phase, and the PSF measurement is simulated as the Fourier transform of the product of the pupil amplitude and the complex exponential of the phase plus the bias

$$PSF_n = FT\left[Ae^{i\times(\varphi+b_n)}\right], \qquad n: 0,1,2 \tag{4-5}$$

, where FT is the Fourier transform operator. To imitate a real scenario the results are normalised to a defined number of photons per detected image and Poisson noise is added to each pixel in the result.



(a) (b) (c)

Figure 4-6. Simulated measurement of three initial PSFs corresponding to the same pupil phase with distinct added biases of positive defocus (a), astigmatism (b) and minus defocus (c).

Three simulated measurements of PSFs corresponding to the aberrated pupil plus biases are shown in Figure 4-6. Compared with Figure 4-5, it can be seen that the PSFs are similar when the biases in the pupil plane are in the same direction (see (b) and (c) sub-images in each figure).

The G-S algorithm with 75 iterations is used to recover the pupil phase from the random initial guess φ_g towards the target φ. The initial pupil field F_0 can be calculated by multiplying the amplitude mask and the exponential term of the phase guess as

$$F_0 = Me^{i\times(\varphi_g)} \tag{4-6}$$

. In one G-S cycle, each of the three biases is added to the initial guess of the pupil phase φ_g, respectively, and the field of the PSFs can be calculated according to the equation 4-5. To make it clear, all the measured PSFs are written in uppercase, while the calculated ones are in lowercase.

$$psf_{g_n} = FT\left[F_0 e^{i\times(b_n)}\right], \; n: 0,1,2 \tag{4-7}$$

The phases of the three estimated psf_{g_n} are calculated using the angle function as $angle(psf_{g_n})$, while the amplitudes of the three psf_{g_n} are then substituted by square-root of the three measured PSF_n (Figure 4-6). The new estimates $\sqrt{PSF_n}e^{i\times angle(psf_{g_n})}$ then undergo inverse Fourier transform to calculate new estimates of the pupil field with three wavefront biases, respectively. To deduce the new estimated pupil field without the impact of the three biases, F_g, these individual estimates are multiplied by the conjugate exponential term of the individual applied bias, and the three results are summed up.

$$F_g = \sum_{n=0}^{2} iFT\left[\sqrt{PSF_n}e^{i\times angle(psf_{g_n})}\right] \times e^{-i\times b_n}, \qquad n: 0,1,2 \tag{4-8}$$

In the pupil plane, instead of a substitution by measurements, the amplitude is smoothed by fitting to a set of 2-dimensional Chebyshev polynomials T_{a_s} from equation 4-19. In the practical simulation, the amplitude values are set to a constant (the 2-D mask M) while the last obtained rms is larger than π since the pupil phase is still changing a lot. (The rms is initially has a large value of over 10). Once the rms change becomes smaller this means that the phase is recovered sufficiently, and the amplitude of F_g is obtained as its absolute value

$$A_g = |F_g| \tag{4-9}$$

and is then fitted to the polynomial expansion.

Despite undergoing 75 iterations of the G-S algorithm, as shown in Figure 4-7, the phase aberration is too large to converge to a correct result. Compared to the target phase φ shown in (a), there are obvious horizontal lines in the unwrapped result φ_{GS} (b), which are caused by phase singularities. The phase is unwrapped along both x and y axes [80], successively, to ensure that phase steps in both directions are removed. In practice, numpy.unwrap is used to adjust the phase values to generate a continuous phase map by removing jumps. However, there is a phase singularity that cannot be eliminated by unwrapping as it corresponds to points where the local phase has a helical variation. The singularities also lead to the speckles in the PSF estimate displayed in (d), even though the target PSF demonstrated in (c) is roughly retrieved. Clearly, the PSF cannot be corrected to a diffraction limit spot only according to just these three initial PSF measurements, as it is difficult to avoid the phase singularities generated in the G-S iteration process.

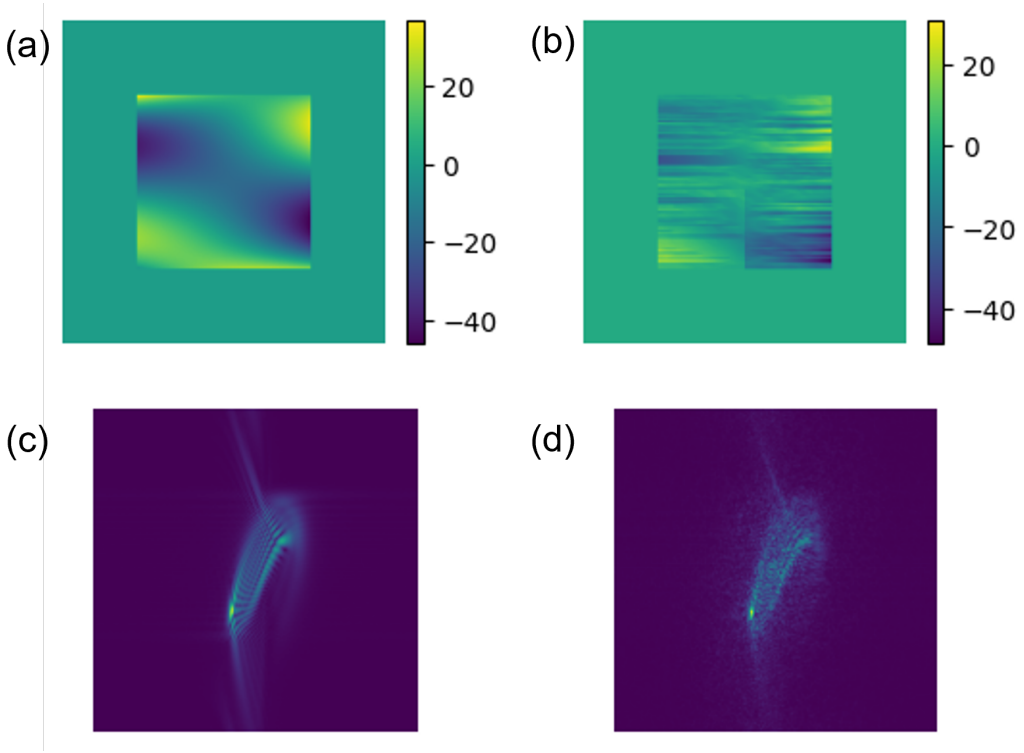


Figure 4-7. Estimating the pupil plane based on PSFs by the G-S algorithm. (a) Target phase φ. (b) Unwrapped phase estimate φ_{GS}. (c) The target PSF. (d) Estimated PSF. All results are acquired after the first 75 G-S iterations.

4.2.3. Smoothing phase and adding feedback via the optical system

To solve the phase singularity problem, the retrieved result from the G-S iterations is smoothed by fitting to polynomial functions. After that, the smooth phase is subtracted from the SLM's phase and used to generate an updated set of measured PSFs that can then be used in a further G-S algorithm step to retrieve the remaining phase aberration. In general, the initial aberration caused by the SLM is significant and cannot converge to a correct estimate merely using the G-S algorithm. Therefore, by applying partial corrections as feedback via the real (or simulated) optical system, the entire aberration can be reduced iteratively to a point where the G-S algorithm produces an estimated result without phase singularities.

To make the phase retrieved by the G-S algorithm continuous, a smooth function is used to fit the phase. The unwrapped phase is fitted over its square aperture to a smooth polynomial expansion, using either a sampled or weighted Chebyshev polynomial expansion estimate. A number (k) of Chebyshev expressions are exploited to describe the phase, and the coefficient for each expression α_k can be calculated as

$$\alpha_k = c_k \sum_{i,j} \varphi_{ij} B_{kij} \qquad 4\text{-}10$$

, where c is the normalisation coefficient of each Chebyshev mode (see the calculations in 4.3) and B is the two-dimensional Chebyshev polynomial. The fitted phase φ_c can be written as

$$\varphi_c = \sum_k \alpha_k B_{kij} \qquad 4\text{-}11$$

As shown in Figure 4-8 (b), the phase φ_c is fitted by the sampled Chebyshev polynomial fit method. The result is smooth compared with the G-S result showed in (a). All the phase singularities are eradicated, and the phase is continuous.

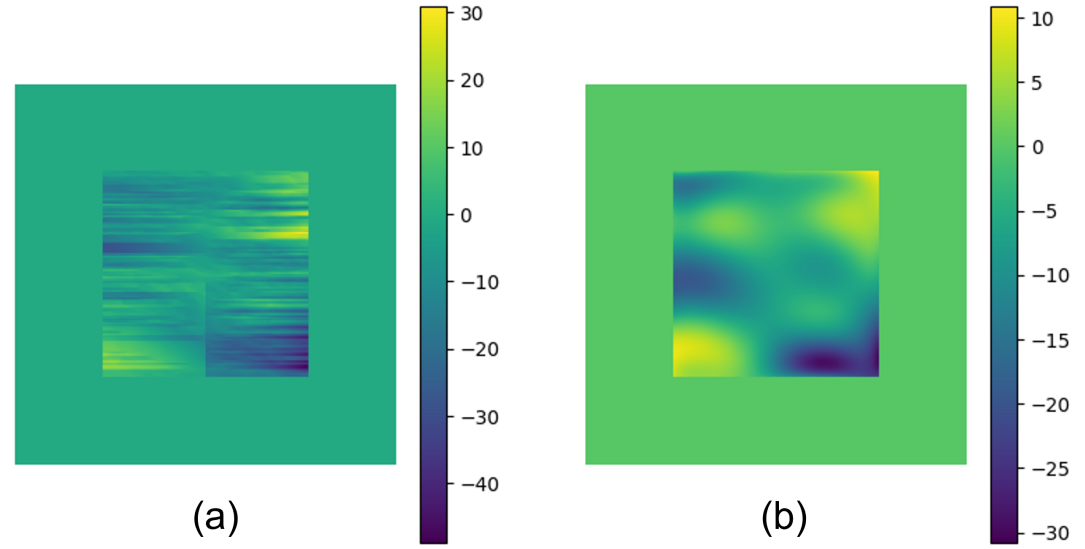


Figure 4-8. Phase smoothing. (a) Unwrapped phase (φ_{GS}) retrieved by the first G-S iteration. (b) Unwrapped phase fitted by Chebyshev polynomials: φ_c.

The fitted phase is then removed from the pupil phase and a new set of PSF_n are measured (or simulated). However, the fitted retrieved phase in one correction iteration only contributes a small part to the aberration. Therefore, this correction is iterated with cycles of G-S phase retrieval to estimate a further refinement to the phase, which is accumulated and deducted from the pupil phase during iterations until the remaining phase estimated phase change becomes small enough. The criterion to finish the iterative phase retrieval is that both current and the last rms values of the estimated phase change are smaller than 0.01 radians.

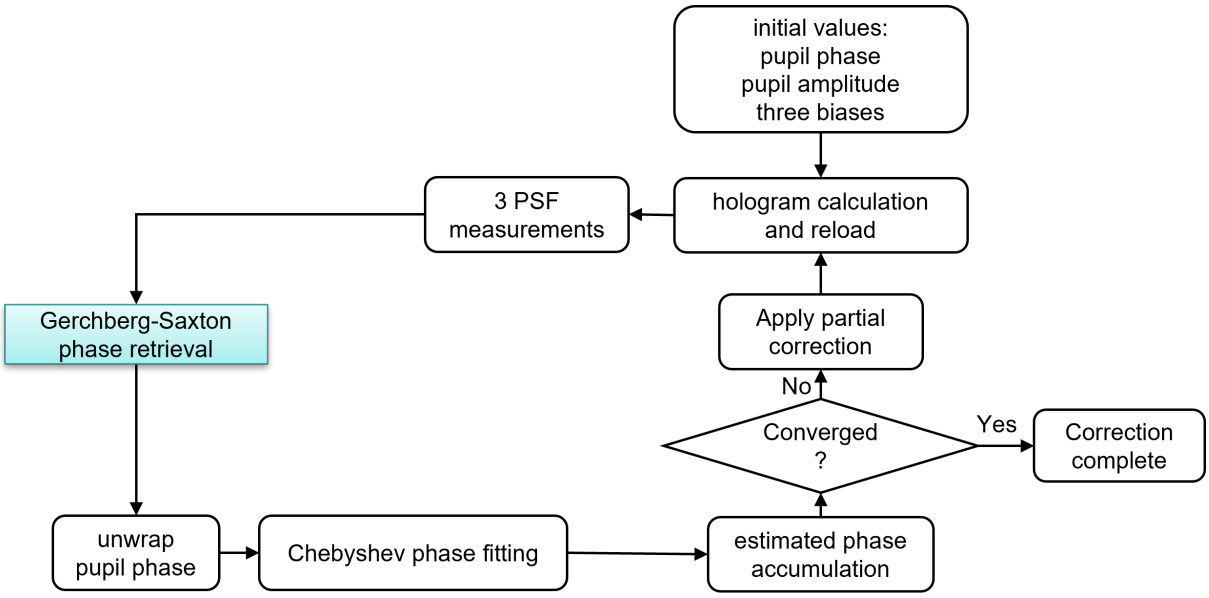


Figure 4-9. Flowchart of the phase recovery.

Figure 4-9 demonstrates the flow chart of the phase correction, including the G-S iteration mentioned above, the phase feedback to update holograms and the rms feedback to finish the process. In practical cases, the result acquired by the G-S algorithm needs to be unwrapped and then smoothed by a polynomial expansion, before the smoothed phase can be holographically removed from the pupil. The hologram is updated by simulating beam interference with the retrieved pupil phase and sent to the SLM. The diffracted orders (PSF_n) will become be less distorted since the phase aberrations are reduced. This can be seen as a partial correction that needs to be repeated because of the large phase aberration cannot be retrieved correctly by the G-S algorithms alone. The hologram is thus updated repeatedly to generated new PSF_n measurements used in the G-S algorithm, and in each correction iteration the accumulated phase corrections applied in the next set of optical measurements. The extent of the partial correction at each iteration can be calculated as rms (root mean square) of the smoothed phase, which is also used as the criterion for completion the value will converge to zero once the correction has converged. In reality convergence is determined once the RMS form two successive correction steps is smaller than a defined value. The Strehl ratio is a commonly used measure of optical system performance, which is defined as the ratio of the peak intensity of the actual PSF in the system to an ideal value, indicating how close the system's imaging quality is to the theoretical maximum. A Strehl ratio of 0.8 is regarded as a benchmark for a diffraction-limited system, roughly corresponding to a wavefront error of about $\frac{\lambda}{13.4}$ RMS (root mean square). Therefore, the criterion value for wavefront RMS coverage in the phase recovery is set to be lower than $\frac{2\pi}{13.4} \approx 0.469$ radians. As is demonstrated below in

section 4.4 and 4.5, the SLM's phase can thus be recovered in only a few correction iterations, and the corresponding diffraction orders can be corrected to close to a diffraction limited result.

4.3. Chebyshev polynomial expansion for a square aperture

The phase recovery algorithm is implemented by fitting at various points to Chebyshev or Zernike polynomial expansions depending on the shape of the physical pupil aperture. There is extensive research describing optical aberrations using Zernike modes [71] [81], however, this section does not specifically explain the generation of these modes. The application of the Zernike approach is presented later.

In the pupil plane, the electric field can be approximated using different approaches and the shape of the pupil can be various. A favourable method is using a Chebyshev polynomial approximation, which suits the square-shaped screen of the SLM in our system since the 2-D ones are defined over $[-1,1]$ both in x and y directions, masking a square in a 2-D plane. Chebyshev polynomials of the first kind $T_n(x)$ can be obtained recursively as functions in one dimension as follows:

$$T_0(x) = 1 \qquad \text{4-12}$$

$$T_1(x) = x$$

$$T_n + 1(x) = 2xT_n(x) - T_{n-1}(x)$$

. The first five Chebyshev polynomials are plotted in Figure 4-10.

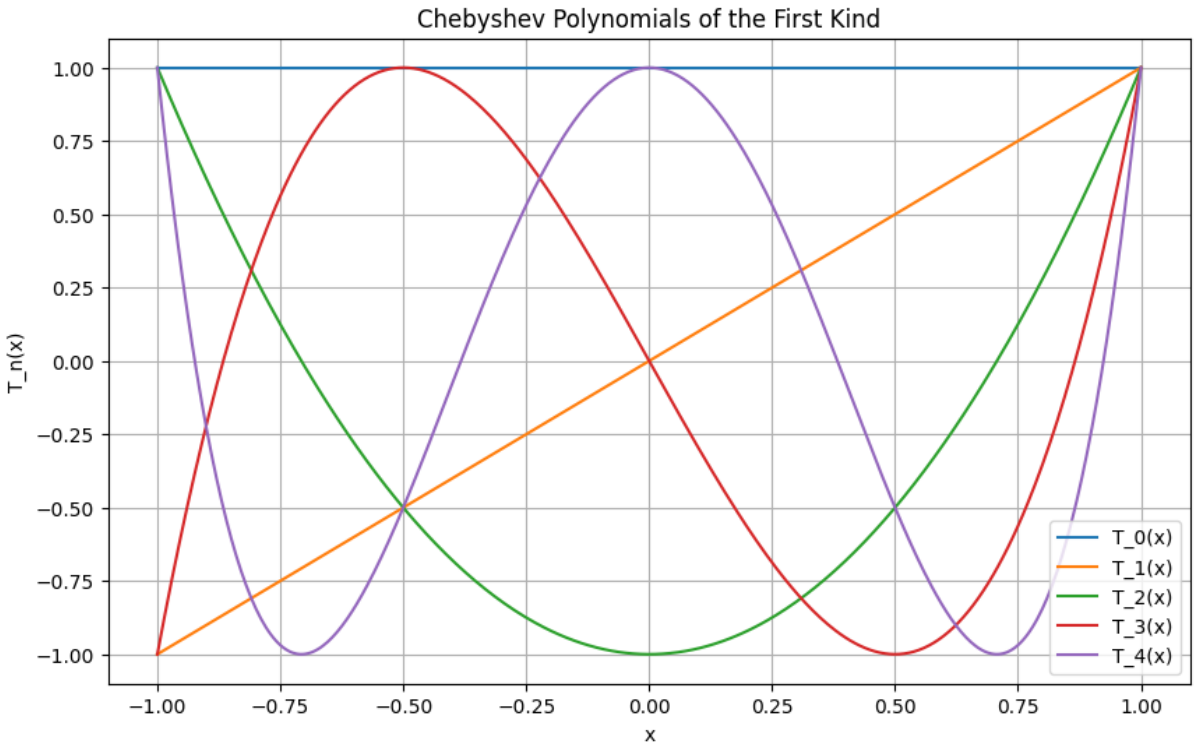


Figure 4-10. Chebyshev polynomials of the first kind.

Orthogonality is an essential property of Chebyshev polynomials. For one-dimensional Chebyshev polynomials of the first kind, orthogonality is defined with respect to a specific weight function over a specific interval. For two general functions $f(x)$ and $g(x)$, if the integral of the product of them and a weight function $w(x)$ over the interval $[a, b]$ equals to zero:

$$\int_a^b f(x)g(x)w(x)dx = 0 \tag{4-13}$$

, they are orthogonal over the rage from a to b. The orthogonality of the Chebyshev polynomials of the first kind $T_n(x)$ is defined over $[-1,1]$ with the weight function $w(x) = \frac{1}{\sqrt{1-x^2}}$. As $n \neq m$,

$$\int_{-1}^{1} T_n(x)T_m(x)\frac{1}{\sqrt{1-x^2}}dx = 0 \tag{4-14}$$

. Due to the orthogonality, the Chebyshev polynomials can be exploited to expand a function with a high computation efficiency and low errors. Compared to using non-orthogonal polynomials, the Chebyshev ones are of higher numerical stability and results tend to converge.

To be more specific, the integral leads to the weighted calculation:

$$\int_{-1}^{1} T_n(x)T_m(x)\frac{dx}{\sqrt{1-x^2}} = \begin{cases} 0 & if\ n \neq m \\ \pi & if\ n = m = 0 \\ \pi/2 & if\ n = m \neq 0 \end{cases} \quad 4\text{-}15$$

. And the discrete form that leads to the sampled calculation:

$$\sum_{k=0}^{N-1} T_n(x_k)T_m(x_k) = \begin{cases} 0 & if\ n \neq m \\ N & if\ n = m = 0 \\ N/2 & if\ n = m \neq 0 \end{cases} \quad 4\text{-}16$$

, where N is any integer greater than $max(n, m)$ and x_k are given by the zeros of $T_N(x)$:

$$x_k = \cos\left(\pi\frac{2k+1}{2N}\right) \quad 4\text{-}17$$

. Also, that two-dimensional Chebyshev polynomials are defined simply as:

$$T_{nm}(x, y) = T_n(x)T_m(y) \quad 4\text{-}18$$

Two choices to perform Chebyshev polynomial fit are possible, a sampled method and a weighted method are explicitly described below.

4.3.1. Sampled Chebyshev polynomials

In this method the discrete orthogonality relation of Chebyshev polynomials is used to calculate a polynomial approximation to an arbitrary function. Here we only need to sample the arbitrary function at a finite set of points to obtain an accurate and efficient approximation. To obtain a polynomial expansion of up to (N) terms in each direction we need to sample at grid of $(M \times M)$ points where $M \geq N$ and the points are given by the nodal points (zero-crossings) of T_M Here we choose an expansion of up to 7 terms in each direction since higher order modes provide only a small contribution to the overall aberration [82], giving a basis of 49 two-dimensional Chebyshev polynomial functions. The number of sampling points is chosen as 8×8, which is larger than the minimum of 7×7. The number of sampling points is set as eight for the simplicity and a reasonable calculation speed as they are enough to perform the estimation for a 7-term expansion, which also determines the grid size of a 2-D polynomial expression. While in one dimension, the zeros where eight polynomials intersect the x axis are also called as nodal points. Chebyshev polynomial functions sample at points intersecting the x or y axes to approximate the functions of the pupil phase (and amplitude). The highest degree of

Chebyshev polynomials is set as n_c as seven, and it should be borne in mind that the degree is distinguishable with the zeros number.

The corresponding number of total Chebyshev polynomials is the square of the max degree because the polynomials used here are two-dimensional. Thanks to Chebyshev functions in Python (numpy.polynomial.chebyshev.chebval2d), the 49 ($n_c \times n_c$) basic Chebyshev expressions corresponding to the seven orders can be calculated as a B_{ij}, where i and j is the index along the vertical and the horizontal axis, respectively. Values of all pixels can be written in a $49 \times 8 \times 8$ matrix $T_{kij}(P_x, P_y)$, where i and j is the vertical and horizontal index, respectively, and is taken from 0 to 6 since the max order is 7. As demonstrated in Figure 4-11, each Chebyshev expression is a 2-D matrix with a shape as 8×8.

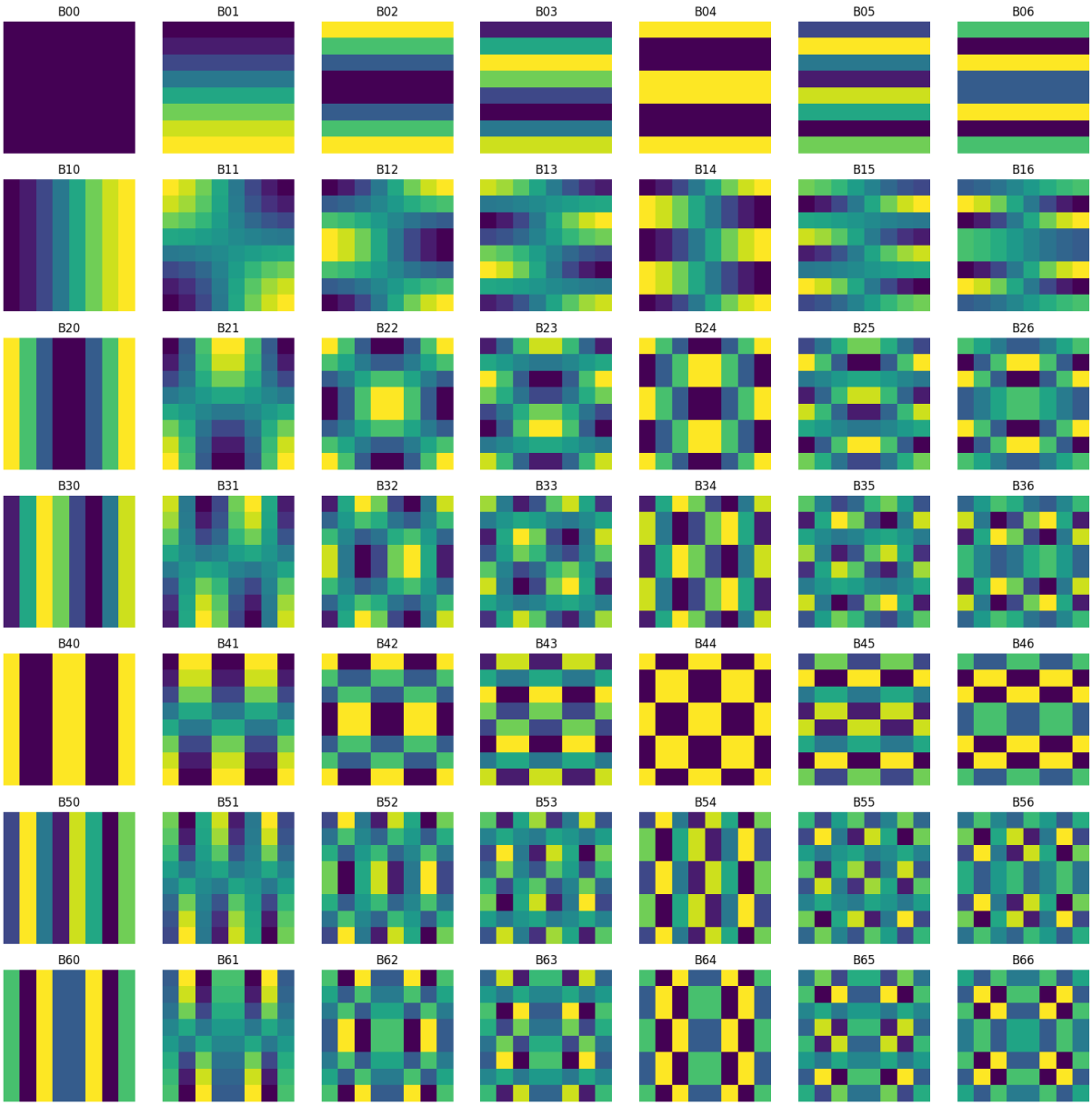


Figure 4-11. 2-D Chebyshev polynomials B_{ij} sampling at nodal points. The subscripts denote the location of the polynomial values in the matrix plot.

Based on the 49 polynomials, the Chebyshev polynomials for fitting the pupil phase T_{p_s} and smoothing the amplitude T_{a_s} can be taken as two sub-matrices, where the subscript s annotate the 'sampling' method. The degree of T_{p_s} is defined as seven which means that all base polynomials are included as more polynomials are required to estimate the pupil phase. While in terms of the amplitude, the order of T_{a_s} is limited to a number n_a as four to simulate the amplitude of a Gaussian illumination. Therefore, the first four by four sub-matrix is taken as T_{a_s}. The number of Chebyshev expressions applied for the phase and the amplitude is denoted as n_c, which is 49 and 16 of the two cases, respectively.

$$T_{p_s} = \begin{bmatrix} T_{00} & T_{01} & \cdots & T_{06} \\ T_{10} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & T_{56} \\ T_{60} & \cdots & T_{65} & T_{66} \end{bmatrix}, \qquad T_{a_s} = \begin{bmatrix} T_{00} & T_{01} & T_{02} & T_{03} \\ T_{10} & T_{11} & T_{12} & T_{13} \\ T_{20} & T_{21} & T_{22} & T_{23} \\ T_{30} & T_{31} & T_{32} & T_{33} \end{bmatrix} \quad \text{4-19}$$

However, the orthogonality relationship of the Chebyshev polynomials requires to be verified to make sure that each polynomial is independent to others. Chebyshev polynomials are of a discrete orthogonality that the sum of the product of two different polynomials equals zero, and the orthogonality O_{km} can be written as

$$O_{klmn} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} T_{kl}(x_i, y_j) T_{mn}(x_i, y_j) = \begin{cases} 0 & if\ k \neq m\ or\ l \neq n \\ N^2 & if\ k = m = 0\ and\ l = n = 0 \\ N^2/2 & if\ k = m \neq 0\ and\ l = n = 0 \\ N^2/2 & if\ k = m = 0\ and\ l = n \neq 0 \\ N^2/2 & if\ k = m \neq 0\ and\ l = n \neq 0 \end{cases} \quad \text{4-20}$$

, where the k (or m) is the index of a Chebyshev polynomial and depends on the square of the highest degree number. For the Chebyshev matrix of phase T_{p_s}, the k (or m) is 0, 1, 2, … 48 while for the amplitude polynomials the range is from 0 to 15. As shown in Figure 4-12 (a), the calculation in equation 4-20 is performed to all polynomials of T_{p_s}, regardless the values of k and l, and the results are plotted as a 49 by 49 matrix. Apart from values at positions where k equals l, each summation of multiplication products of two 2-D Chebyshev polynomials is zero, demonstrating the discrete orthogonality relation. Similarly, the calculation is applied to T_{a_s}, and the 16×16 matrix plot is illustrated in Figure 4-12 (c).

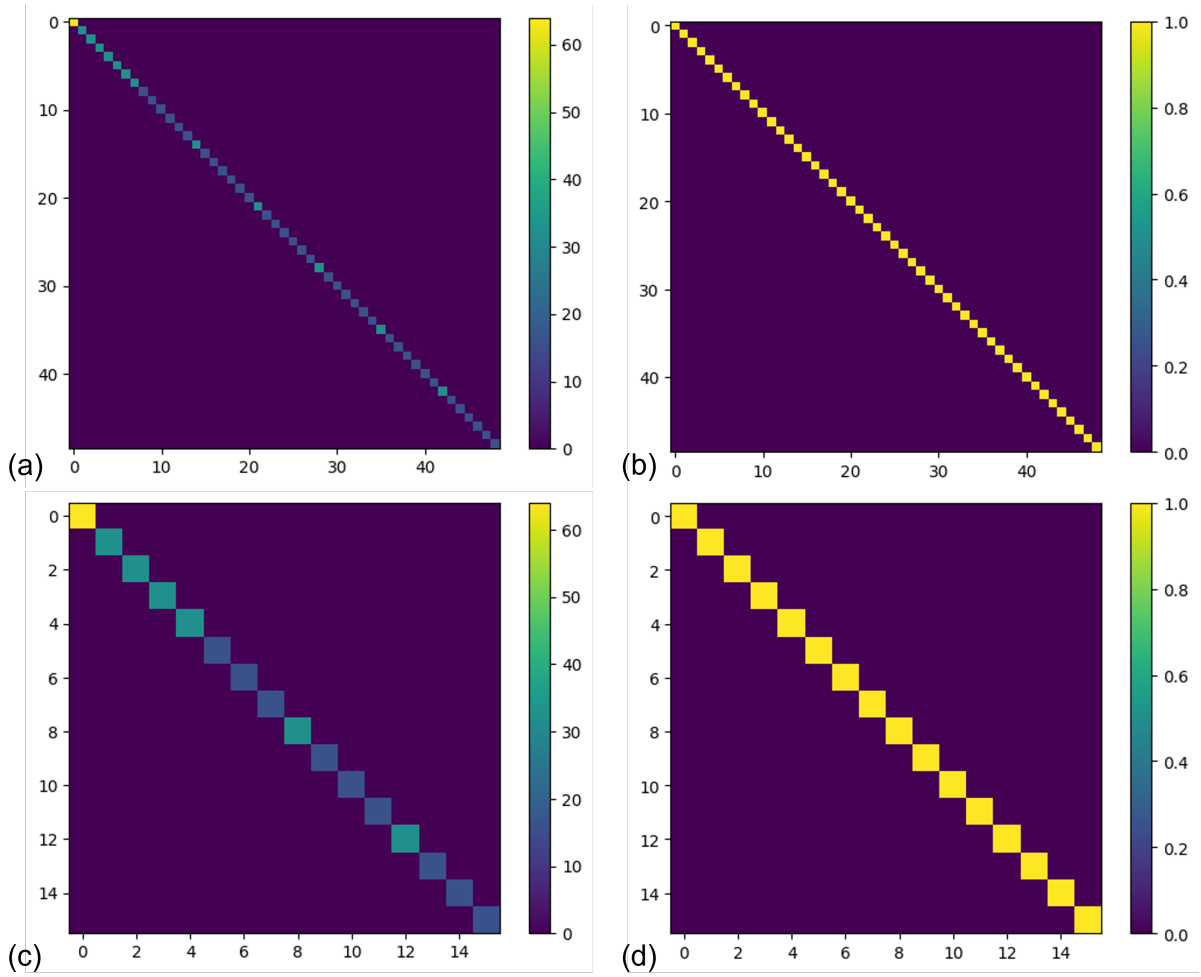


Figure 4-12. Matrix plots of the orthogonality relation and the normalisation of Chebyshev polynomials approximating the phase and the amplitude of the pupil plane. (a) The discrete orthogonality of the phase ones. (b) Sums of the multiplying products of 2 polynomials and the normalisation values. The orthogonality matrix of Chebyshev polynomials for fitting the pupil amplitude without (c) and with the normalised coefficients (d).

Although both the two matrices of polynomials demonstrate an orthogonal relationship, the orthogonality of Chebyshev polynomials usually needs to be normalised to make the phase and fittings easier. As obviously shown in Figure 4-12 (a) and (c), the orthogonality value changes from 0 to 64, which is required to be normalised as the different weight affects the proportion of each Chebyshev polynomial contributing the pupil's phase or amplitude. In fact, the fitting using polynomials without the normalisation cannot converge at all. Since the normalisation values depend on the particular Chebyshev mode, an array of normalisation coefficients for each mode required to be calculated. A normalisation coefficient can be calculated as the inverse of $O_{k_1 k_2}$ when k_1 and k_2 is the same, which is actually one of diagonal values of the matrix $O_{k_1 k_2}$. Each coefficient c_k is also the inverse of the sum of squares of elements in one polynomial, and is calculated as:

$$c_{s_k} = \frac{1}{\sum T_{kij}(P_x, P_y) T_{kij}(P_x, P_y)} \qquad 4\text{-}21$$

$$i \in \mathbb{Z} \,|0 \le i \le 7, \qquad j \in \mathbb{Z} \,|0 \le j \le 7$$

, where s denotes the sampled method, and $\mathbb{Z}$ represents the group of integers. According to the equation 4-21, the coefficients for the amplitude and the phase fitting are denoted by $c_{a_{s_k}}$ and $c_{p_{s_k}}$, respectively, and the indices are over different ranges as

$$k \in \mathbb{Z} \,|0 \le k < 49 \;\; for \; c_{a_{s_k}}, \qquad k \in \mathbb{Z} \,|0 \le k \le 16 \;\; for \; c_{p_{s_k}} \qquad 4\text{-}22$$

. All normalisation values consist of an array, and the corresponding normalisation array for the phase and the amplitude includes 49 and 16 values, respectively. Given a normalisation array, the orthogonality matrix in equation 4-20 can be normalised as

$$\hat{O}_{km} = \sum_{i,j} c_{s_k} T_{kij}(P_x, P_y) T_{mij}(P_x, P_y) \qquad 4\text{-}23$$

. With the normalisation, the orthogonality matrix for the phase and the amplitude approximation is solely of the value as zero or one, which is illustrated in Figure 4-12 (b) and (d), respectively.

The method of sampling Chebyshev polynomials at nodes is accurate and efficient since the polynomials are only calculated at a few points. When the Chebyshev expressions yielded in this way are used to fit a phase (or an amplitude), however, the phase values over the whole plane need to be interpolated only at the nodal points.

4.3.2. Weighted Chebyshev polynomials

Therefore, another approach to define the Chebyshev polynomials is introduced, which is based on the continuous orthogonality relation and a weight function. However, the polynomials are still calculated at limited coordinates (pixels) to imitate the continuous function. The continuous orthogonality relation between Chebyshev polynomials is based on the integral of the product of two different Chebyshev polynomials over the interval $[-1, 1]$ with a weight function and is given by the following equation.

$$\int_{-1}^{1} T_a(x)T_b(x)\frac{dx}{\sqrt{1-x^2}} = \begin{cases} \pi & if\ a = b = 0 \\ \frac{\pi}{2} & if\ a = b \neq 0 \\ 0 & if\ a \neq b \end{cases} \quad \text{4-24}$$

Suppose we sample the range $[-1, 1]$ at m evenly spaced points separated by $\frac{2}{m}$. The sample points x_i are calculated on a grid where:

$$x_i = -1 + \left(i + \frac{1}{2}\right) * \frac{2}{m} \quad , i = 0..m-1 \quad \text{4-25}$$

And the integral is approximated by

$$\int_{-1}^{1} T_a(x)T_b(x)\frac{dx}{\sqrt{1-x^2}} \approx \frac{2}{m}\sum_{x_i}\frac{T_a(x_i)T_b(x_i)}{\sqrt{1-x_i^2}} = \frac{2}{m}\sum_{x_i} w_i T_a(x_i)T_b(x_i) \quad \text{4-26}$$

Note that the first sample point is at $x_0 = -1 + \frac{1}{m}$ and the last is at $x_{m-1} = 1 - \frac{1}{m}$ so that the weight function w_i avoids the infinities at ± 1. However, there is still a significant error in the summation, particularly because the integral is poorly represented by the value of the summation terms at these end points.

Based on the coordinates defined above, the weight function can be calculated. However, it's obvious that the function goes towards infinity at its end elements x_0 and x_{m-1} which makes the approximation of the integral by simply summing pixel values inaccurate. To find a correction to the weight function for the end elements, two simple Chebyshev expressions are used as an example. Considering the explicit formulas of the first three one-dimensional Chebyshev polynomials as

$$T_0(x) = 1, \qquad T_1(x) = x, \qquad T_2(x) = 2x^2 - 1 \quad \text{4-27}$$

, T_0 and T_2 are chosen to empirically determine a correction of the weight value at the end points so that the sum over the product of these two polynomials computes correctly as zero. Given the integral (summation) in equation 4-26, the adjusted end weights w_0' and w_{m-1}' can be considered as the same and calculated from:

$$w_0' = w_{m-1}' = w_o - \frac{\left(\frac{2}{m}\sum w_0 T_0(x_0)T_2(x_0)\right)}{T_0(x_0)T_2(x_0)} \quad \text{4-28}$$

$$2$$

, assuming w_i is zero when $i \neq 0 \; or \; m-1$. The discrete summation replacing w_0 with w_0' and w_{m-1} with w_{m-1}' will now result in the correct summation of zero for the pair of polynomials T_0 and T_2 These modified weight values for a 2-dimensional calculation of Chebyshev polynomials orthogonality are presented in Figure 4-13. The values of the pixels around the four endpoints (vertices of the mask) are still much higher than the other weights even though they are modified for the continuous orthogonality calculation.

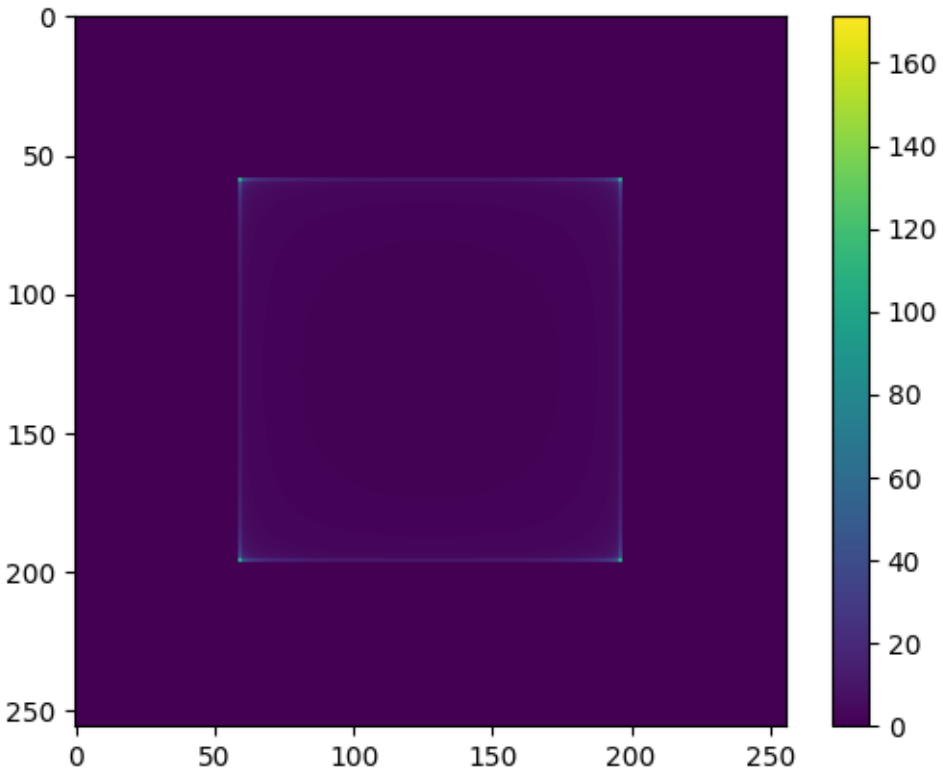


Figure 4-13. The weighting values over the pupil plane. The four vertices of the square are endpoints of the mask.

The individual Chebyshev polynomials calculated across the same region as in Figure 4-13, are illustrated in Figure 4-14.

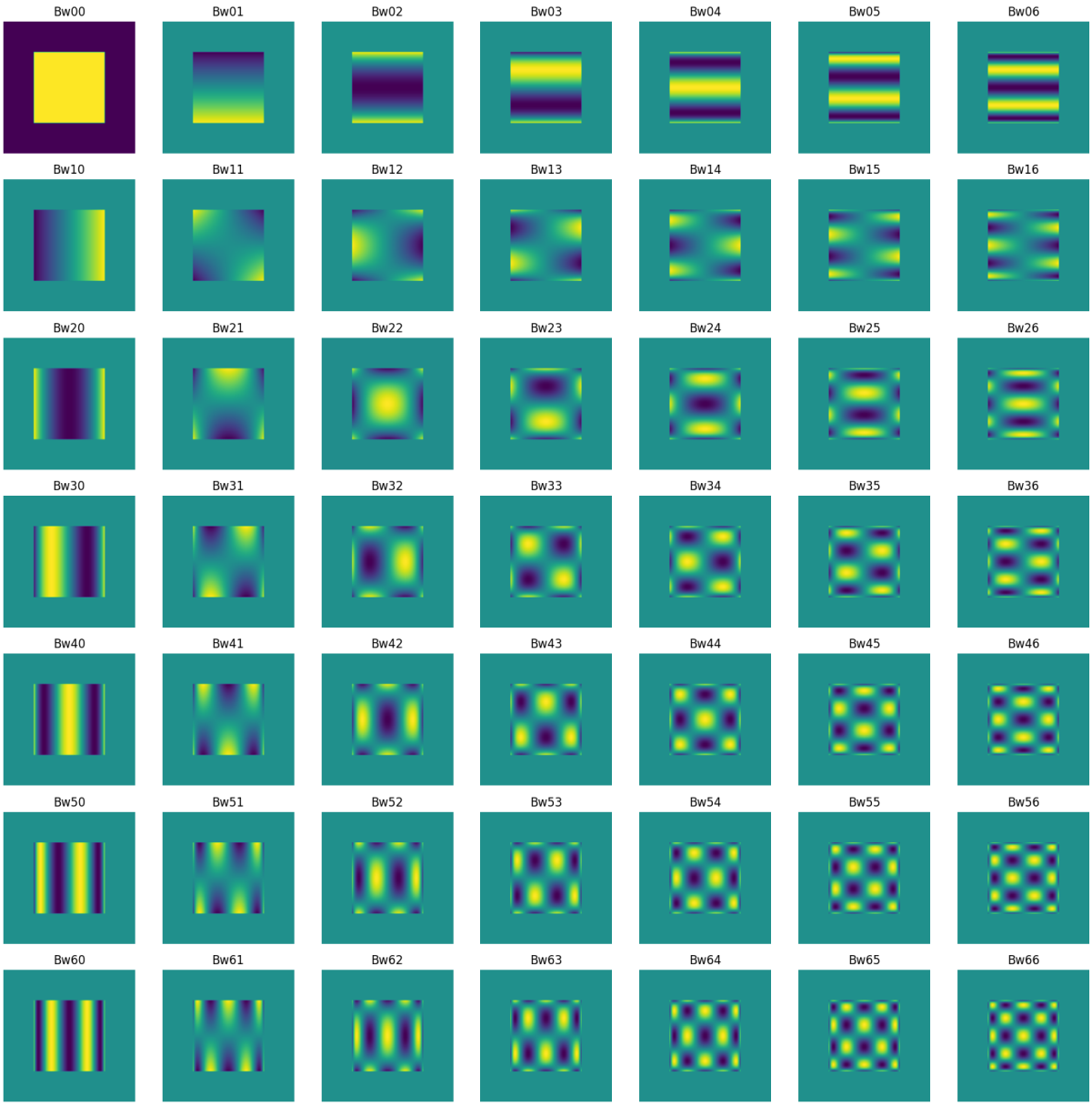


Figure 4-14. Two-dimensional Chebyshev polynomials $B_{w_{ij}}$ over the pupil plane with a 2-D mask.

The orthogonality of the weighted Chebyshev polynomials can be then examined and plotted with normalisation in with and without normalisation, in an approach that is similar to the sampling Chebyshev method. Referring to equation 4-26 and considering the weights calculated in equation 4-28. The matrix plot of the normalised orthogonality matrices of the phase polynomials is showed in Figure 4-15 using uncorrected (a) and corrected (b) weights, respectively.

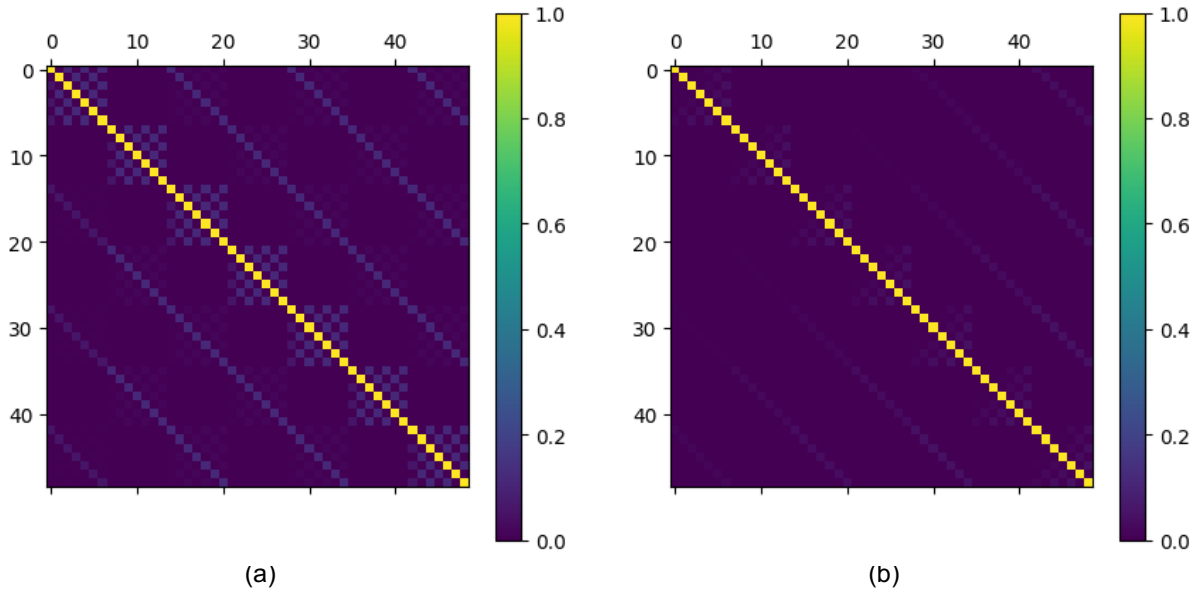


Figure 4-15. The orthogonal relation of Chebyshev polynomials obtained with weighted summation. The normalised orthogonality of the Chebyshev expressions is shown both before (a) and after (b) correction of the weighting factor at the end points.

In the plots, apart from the diagonal values, there are still some values are not equal to zero, which demonstrates the deficiencies of this weighted Chebyshev calculation method. Even though the correction reduces the off-axis terms between 0^{th} and 2^{nd} order polynomials to zero, it does not completely suppress other off-diagonal terms. In the uncorrected version the maximum off-diagonal term is 0.117 radians and the mean is 0.012 radians, while with corrected weights, the maximum off-diagonal term is 0.044 radians and the mean is 0.0022 radians.

4.4. Simulation results

4.4.1. Comparison between two Chebyshev fitting methods

By combining the G-S algorithm and the feedback to optics with the sampled Chebyshev expansions, the SLM's phase can thus be recovered in only a few correction iterations, and the corresponding diffraction orders can be corrected. Some intermediate results are shown in Figure 4-16. As shown in (a), the 0^{th} PSF calculated in the G-S algorithm is close to the result for a PSF showing only with the zeroth bias. Figure 4-5 (d), shows the estimated G-S correction and that there hardly is any residual aberration in the pupil field. Compared to the simulated PSFs without the SLM aberration in Figure 4-5, the simulation of the measurements showed in (j) to (l) as a result of the complete correction have converged to the case of small

residual aberration. The total accumulated phase in (i), is similar to the original applied aberration shown in (f) up to a constant phase shift, and the estimated pupil intensity in (d) can be seen as a Gaussian light source in the square pupil. In addition, as demonstrated in (e), the rms of the retrieved phase converges to near zero at the final iteration, indicating that the phase used as the final partial correction is near uniform after eight correction iterations. The residual aberration is showed in the estimated phase (b), the unwrapped counterpart, showing no singularities (g) and then smoothed by Chebyshev expansions (h). This final phase correction and has a residual value of 0.0032 radians. (The criterion is set to be 0.01 radians.)

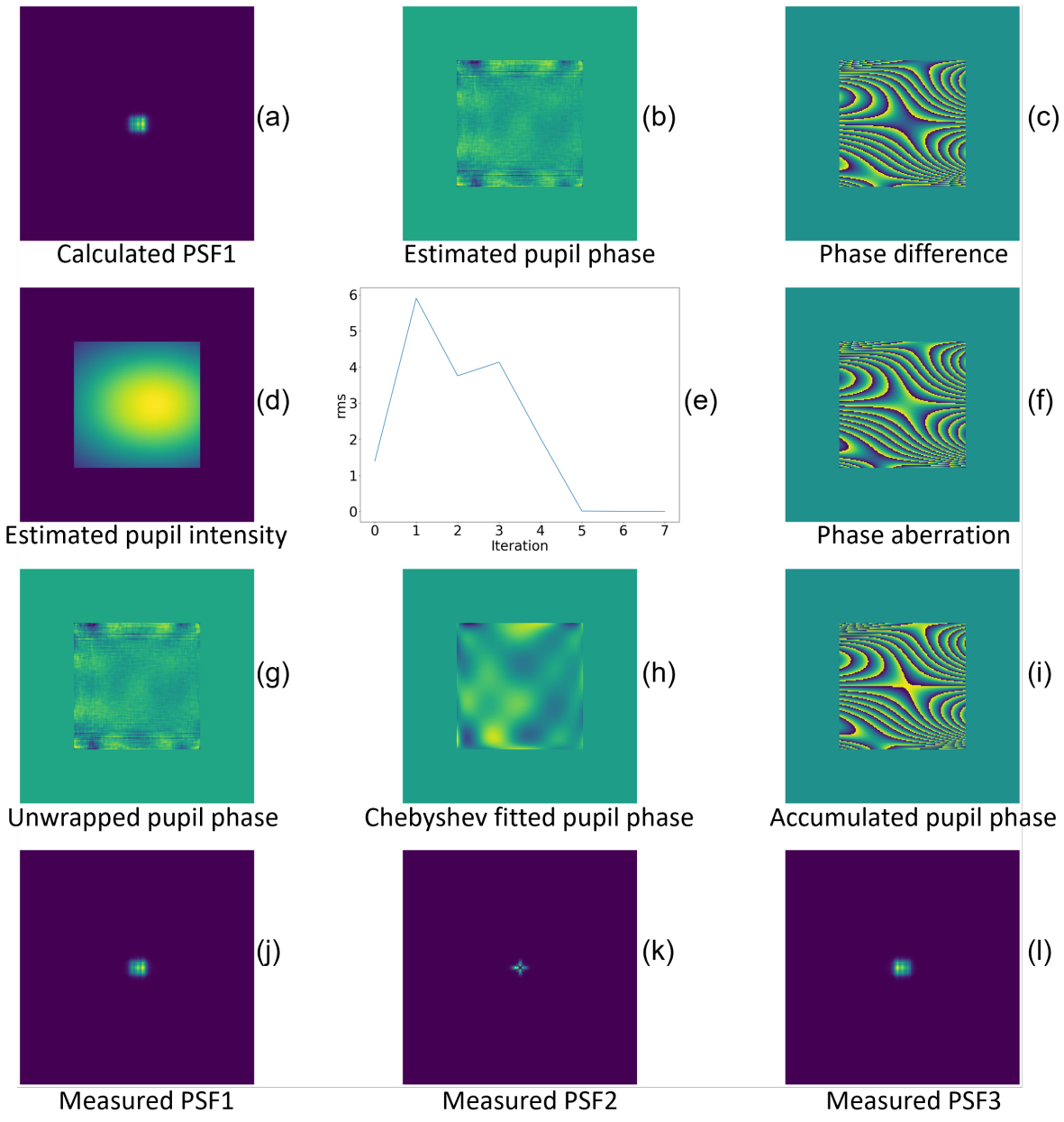


Figure 4-16. Phase recovery with sampled Chebyshev fitting. (a) The calculated PSF with the aberration and 0th bias: psf_{GS_0}. (b) The estimated G-S pupil phase: φ_{GS}. (c) φ_{GS} minus the actual phase aberration Φ. (d) The

estimated pupil amplitude. (e) The rms of the current fitted phase φ_c. (f) The wrapped simulated phase aberration: Φ. (g) Unwrapped φ_{GS}. (h) The chebyshev fitted phase: φ_c. (i) The wrapped accumulated phase. (j – l) The simulated measurments of corrected PSFs with the 0^{th}, 1^{st} and 2^{nd} bias, respectively. Phases in (c), (f) and (i) are wrapped to cover a range from 0 to 2π by taking the angle function to present more straightforward images.

From left to right, Figure 4-17 shows the progression of a typical correction.

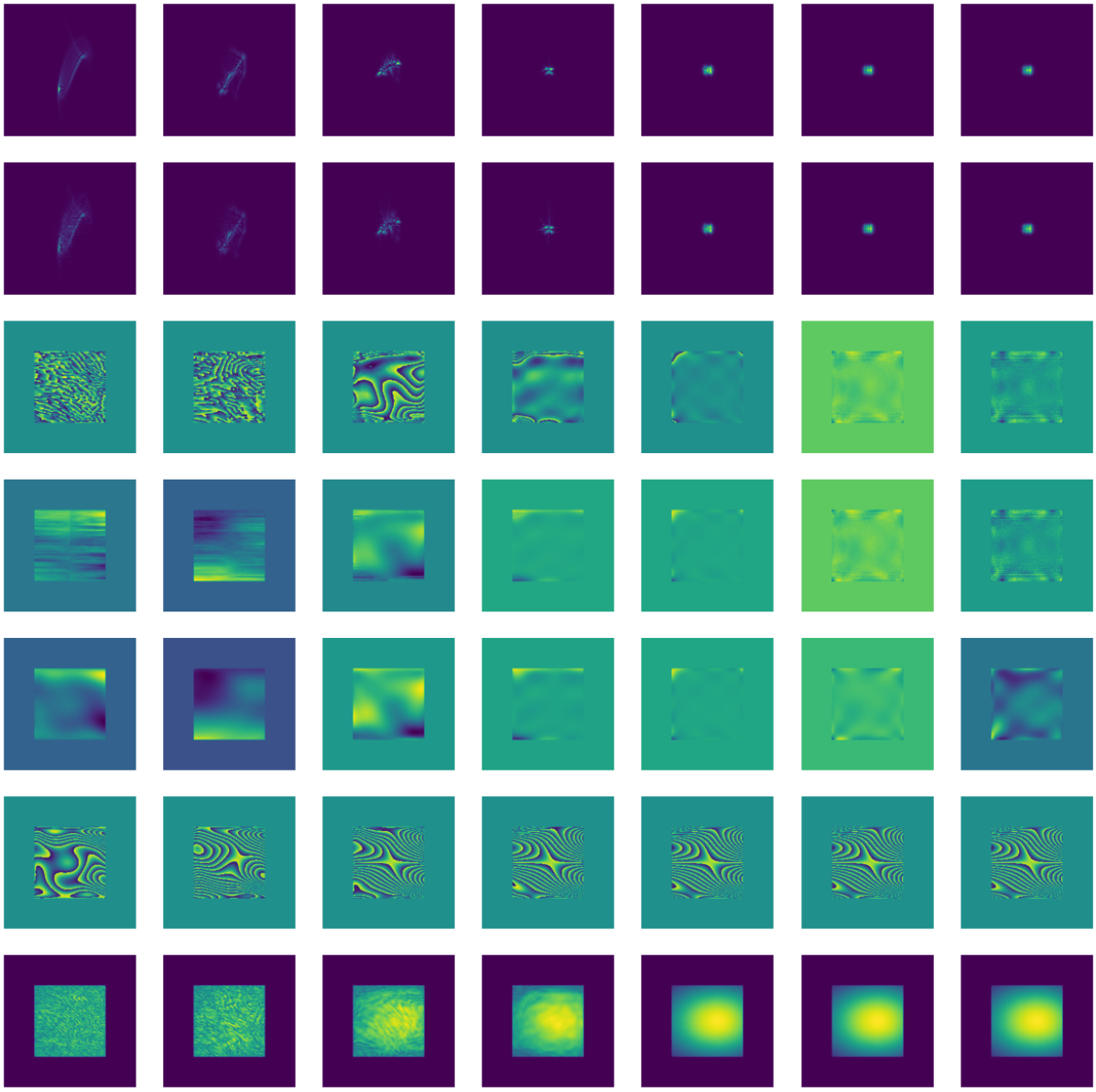


Figure 4-17. The progression of a typical phase recovery. The row from top to bottom shows the calculated PSFs, simulated PSF measurements, estimated phases, unwrapped phase estimations, Chebyshev-fitted phases, accumulated phases, and the pupil intensities, respectively.

The simulation can be repeated by fitting the amplitude and the phase of the pupil using the weighted Chebyshev polynomials. As described in 4.3.2, the weighted Chebyshev expressions can be used without interpolation operations since they include each pixel in the pupil, although some values in the orthogonality matrix are non-zero off-axis.

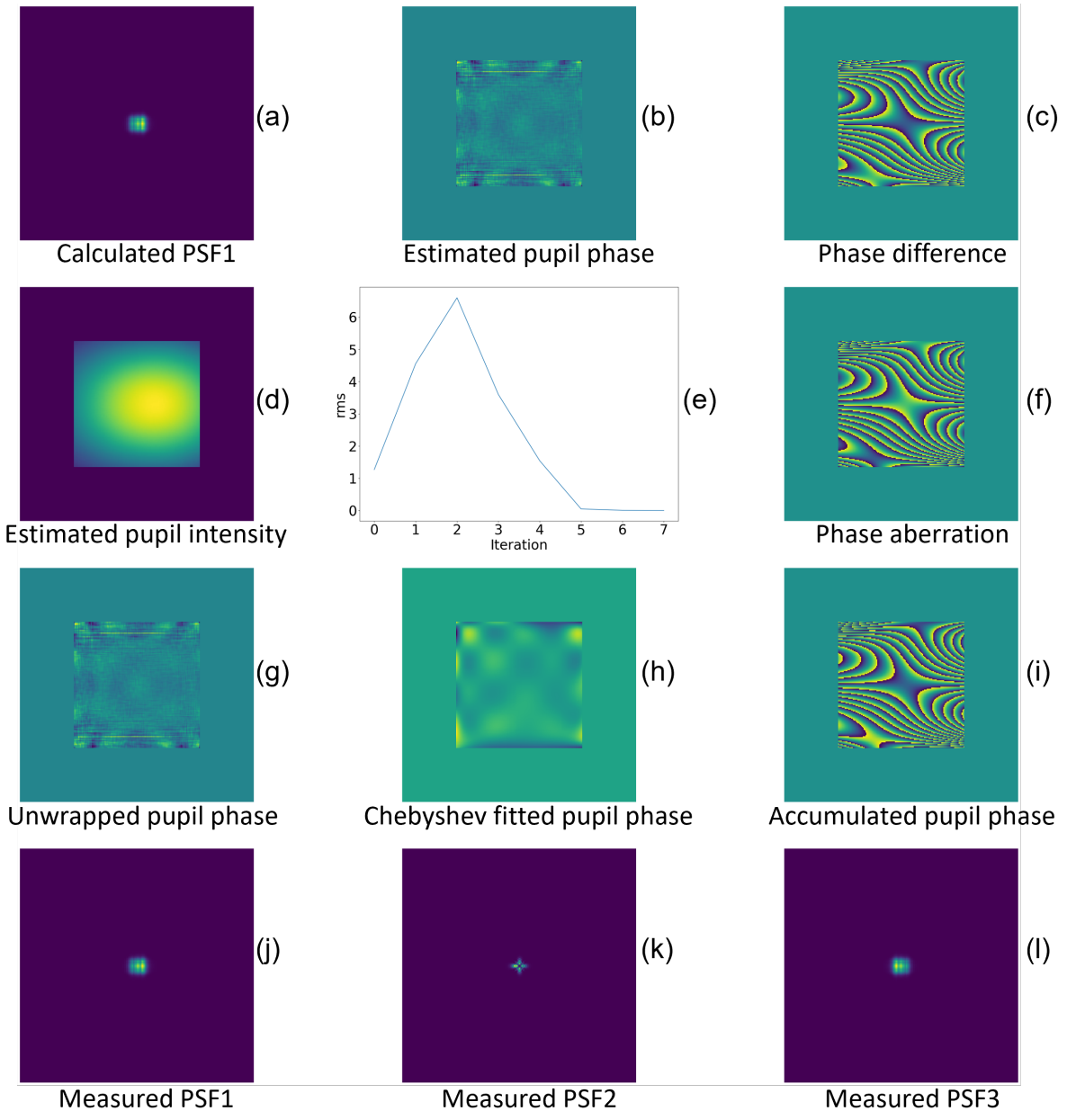


Figure 4-18. Phase correciton with the weighted Chebyshev approach. (a) The calculated PSF with the aberration and 0^{th} bias. (b) The estimated G-S pupil phase. (c) φ_{GS} minus the actual phase aberration Φ. (d) The pupil amplitude estimate. (e) The rms of the current fitted phase φ_c. (f) The wrapped simulated phase aberration: Φ. (g) Unwrapped φ_{GS}. (h) The chebyshev fitted phase: φ_c. (i) The wrapped accumulated phase. (j – l) The simulated measurments of corrected PSFs with the 0^{th}, 1^{st} and 2^{nd} bias, respectively. Phases in (c), (f) and (i) are wrapped to cover from 0 to 2π by taking the angle function to present more straightforward images.

As shown in Figure 4-18, the phase aberration caused by the SLM can also be effectively mitigated by the method using the weighted Chebyshev polynomials. After a few correction iterations, the acquired phases in (b), (g) and (h) are uniform enough since the pupil phase is significantly recovered, and the PSFs in (a), (j), (k) and (l) hence approach the cases without the SLM aberration. The pupil amplitude showed in (d) is close to a gaussian illumination. Although the accumulated phase (i) and the ground truth (f) are not identical, the rms of the partial phase correction presented in (e) reduces to a small value of about 0.0048 radians in the final iteration.

As displayed in Figure 4-19 (a), the simulated SLM aberration leads to an extented PSF, however, which can be recovered back to a diffraction limited spot by applying the phase recovery method. The PSFs showed in (b) and (c) are generated by the algorithm with the sampled and the weighted Chebyshev polynomials, respectively. Although the calculated weighting values cause the imperfect orthogonality of the weighted Chebyshev expressions mentioned in 4.3.2, in the simulation, the SLM aberration can still be practically measured and recovered by the weighted method. Both two treatements lead to reliable estimations of the electric field in the pupil plane.

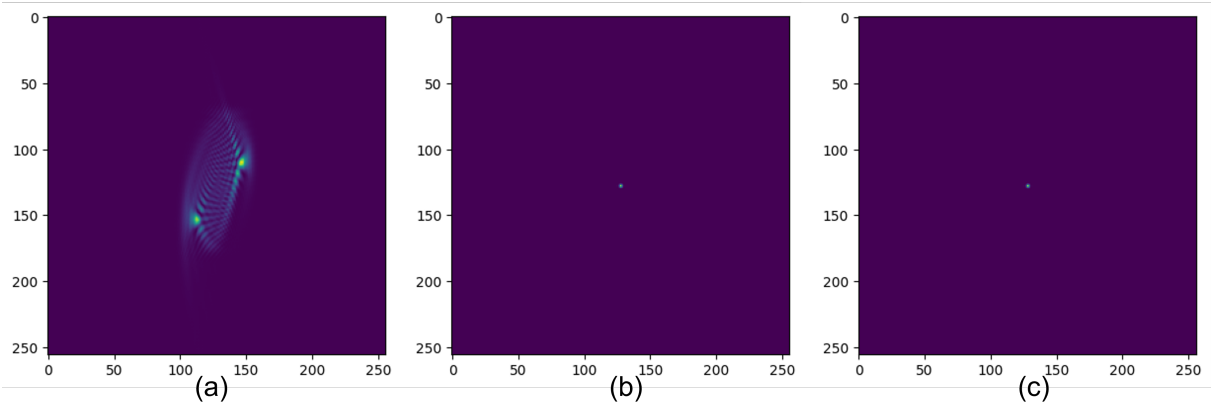


Figure 4-19. PSFs before (a) and after the phase recovery with the sampled (b) and the weighted (c) Chebyshev fittings.

4.4.2. Applying Zernike fitting for circular aperture

The measurement and recovery of the phase aberration can also be realised with smoothing the phase over a circular aperture using Zernike polynomials ($Z_0^0, Z_1^{-1}, Z_1^1 \ldots$). Zernike polynomials have a similar integral orthogonality relationship to Chebyshev polynomials but

the weighting function is unity and so correction is not required. In the simulation, the pupil M is restricted in a circle rather than a square.

$$M: \sqrt{x^2 + y^2} < 1 \qquad 4\text{-}29$$

As show in Figure 4-20 (a), the phase aberration φ introduced by the surface of the SLM is calculated as a combination of a few low order Zernike polynomials:

$$\varphi = (30 \times Z_2^2 + 2 \times Z_2^{-2} - 5 \times Z_3^1 + 7 \times Z_3^{-1} + 5 \times Z_3^3 - 2 \times Z_3^{-3}) \times M \qquad 4\text{-}30$$

. The random starting phase φ_g also consists of a different summation of a few low order Zernike polynomials:

$$\varphi_g = (-18 \times Z_1^1 + 5 \times Z_2^2 - 7 \times Z_3^1 + 11 \times Z_3^3) \times M \qquad 4\text{-}31$$

, and is presented in (e). The three biases are chosen as positive defocus, zero aberration and negative defocus, which are shown in (b) to (d), along with the resulting simulated PSFs including both aberration and bias displayed in (f) to (h), respectively. The determination of values of biases is empirical.

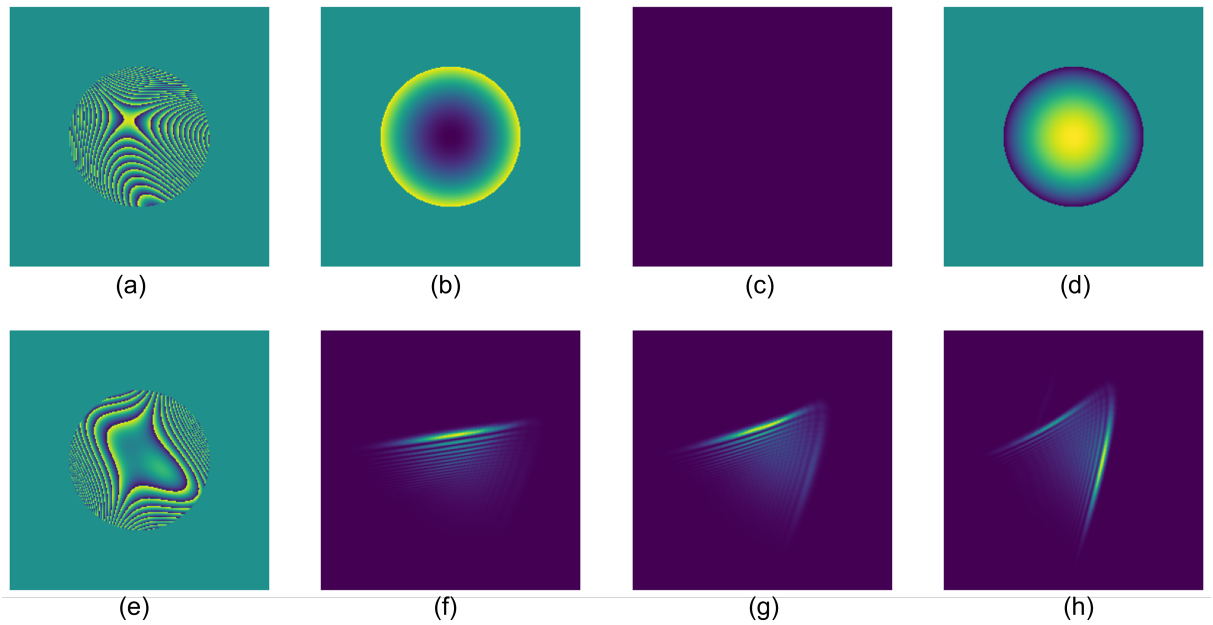


Figure 4-20. Simulated parameters in the phase recovery with the Zernike fitting. (a) φ: The simulated phase aberration at the SLM plane. (b) b_0: $7.5 \times Z_2^0$. (c) b_1: 0. (d) b_2 : $-7.5 \times Z_2^0$. (e) φ_g: The initial guess of the phase aberration. (f) - (h): The PSF corresponding to the aberrated wavefront with biases b_0 - b_2.

The final convereged result is displayed in Figure 4-21, the where the fitting of phase and amplitude estimates in the circular pupil is made using the smooth Zernike polynomials. The criterion is that the RMS must be less than 0.01 radians in two consecutive measurements. The combination of the G-S algorithm and the optical feed back can again retrieve the phase aberration effectively. In terms of the PSF, the calculation of the measured intensity with the 0^{th} bias is showed in (a), and the final simulated measurements with all three biases are shown in (j) to (l). The images of PSFs cannot be over-exposed because they are all simulated, which is why we can choose one of the biases to be zero. In real scenarios, focusing all light on a small spot can lead to an overexposure of the intensity in that spot and inaccurant PSF measurements. Due to the zero bias, the sub-images in Figure 4-20 (g) and in Figure 4-21 (k) are the PSF with the aberration before and after the phase retrieval, respectively. The estimated phase displayed in (b) and (d) is very uniform, although the unwrapped one demonstrated in (h) can highlight the remaining small phase aberation at the pupil plane. While it is difficult to differentiate the simulated phase aberration showed in (f) and the accumulated phase retrieval in (i). However, as displyed in (e), the final rms ends at 0.0076 radians after ten correction iterations, which means that the residual final phase is approximately uniform, and this is also proved by the Gaussian distribution of the pupil amplitude.

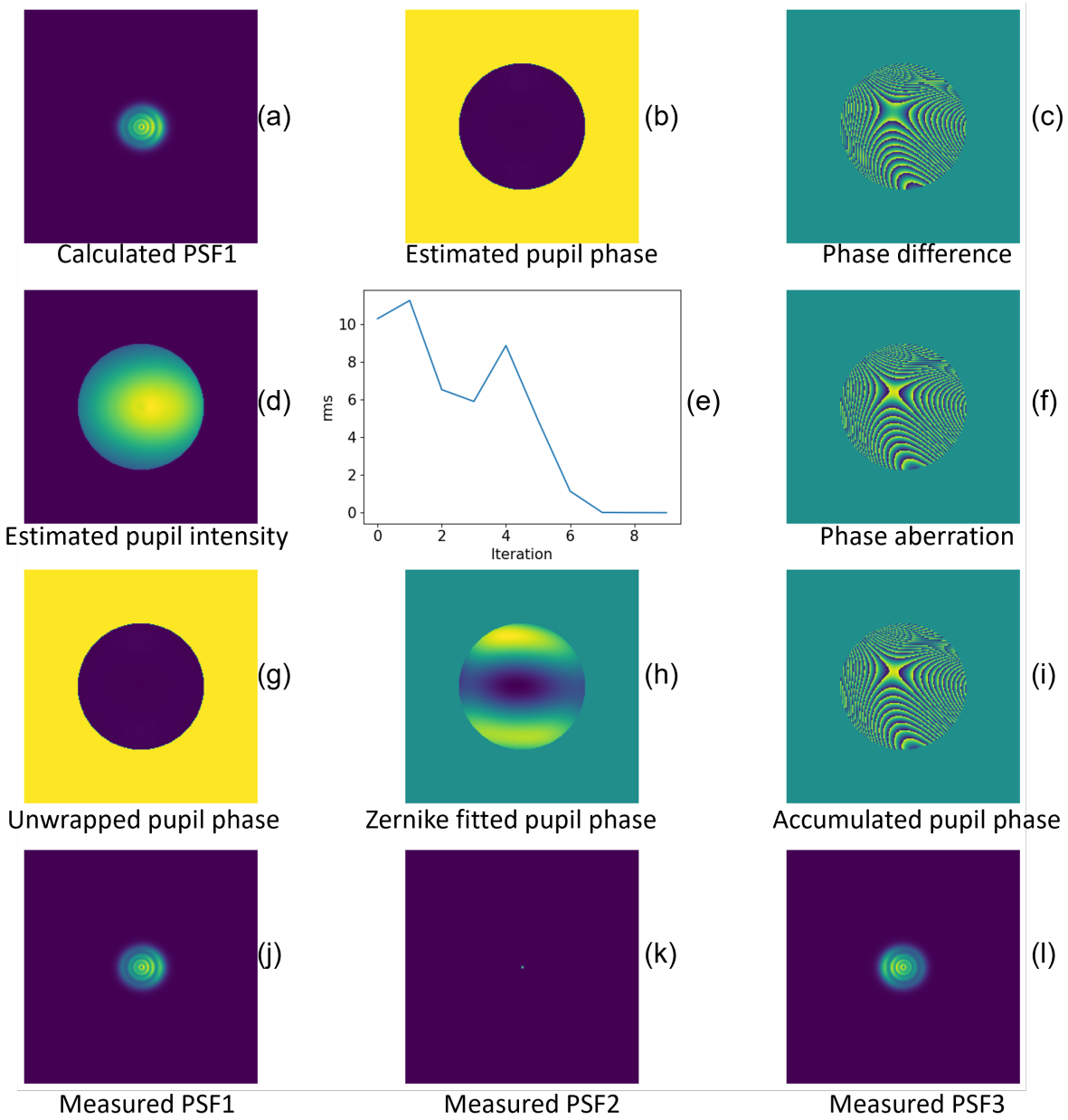


Figure 4-21. Phase recovery with Zernike fitting. (a) The calculated PSF with the aberration and 0th bias: psf_{GS_0}. (b) The estimated G-S pupil phase. (c) φ_{GS} minus the actual phase aberration Φ. (d) The estimated pupil amplitude. (e) The rms of the current fitted phase φ_c. (f) The wrapped simulated phase aberration: Φ. (g) Unwrapped φ_{GS}. (h) The chebyshev fitted phase: φ_c. (i) The wrapped accumulated phase. (j – l) The simulated measurments of corrected PSFs with the 0th, 1st and 2nd bias, respectively. Phases in (c), (f) and (i) are wrapped to present more straightforward images.

4.5. Implementation with a simple optical layout

Apart from the simulations, the phase aberration recovery can be implemented on the simple optical layout shown in Figure 2-1 by combining the Python code and real PSF measurements from the camera.

4.5.1. Writing holograms in the SLM's flash storage on the fly

In implementations, all calculations in the G-S algorithm and the correction iterations are the same as the simulations mentioned above, but the simulated PSFs in the equation 4-5 are substituted by images obtained by the Thorlabs camera. To this end, in each measurement, a hologram of the pupil field is generated and sent to the board of the SLM to diffract the source light into corrected orders (PSFs). In each correction iteration, the hologram is generated by adding the phase bias to the field phase and subtracting the accumulation of retrieved phase. The hologram is then sent to the SLM.

In the SLM's system, the common way to send holograms is to pack images, running orders (how the images are displayed and how the display is activated) and sequences (the period of displaying each image) together as a file called a 'repertoire' and then send the file to the SLM's board. However, the long transfer time required to send a complete repertoire slows down the whole phase recovery process and is partially redundant anyway as resending the running orders and the sequences in each measurement of PSFs is not required. According to the SLM's manual, it takes 15s to send a single 8-bit image, and the transfer time for a complete repertoire including three holograms for three PSFs and other information is longer than 45s. Considering that the several optical iterations would be required, more than five minutes could be consumed in running the phase correction method. In contrast, the G-S process, even with the typical 75 iterations, computes in the order of 1 second.

To speed up the image transmission in each correction iteration, we design a method to safely update the holograms on the fly by directly writing holograms to the flash buffer of the SLM. A repertoire is sent to the SLM, including the settings of the sequences, several running orders and multiple assigned positions of bit-planes in the flash storage buffer, but no actual images. Three bit-plane images are required by each running order to generate the three diffracted PSFs with different bias in each correction iteration. In order to change the images referenced by a particular running order, the flash storage buffer must be updated with the new bit-plane images containing the holograms calculated with the introduced biases and the accumulated phase correction. Flash storage has a limited number of rewrite cycles (100,000 times) so a

new running order is randomly chosen in each correction iteration that accesses data in a new part of the available flash storage.

The SLM comes with a python API providing a number of useful functions including the ability to start a repertoire and change running orders within a given repertoire. It also provides the ability to access flash storage directly and to load new bit-plane data (actually an erase and update process). Following the hologram update, a function (recently provided by the SLM's manufacturer: Forth Dimension Displays) is used to reload the images at the assigned bit-plane positions from the flash buffer, and the holograms can be displayed according to the sequences to generate diffraction orders. Using this process, the optical update step is reduced from 10s of seconds to just a few seconds.

However, the holograms produced from the phase correction method require to be formatted by interleaving before writing into the flash as the display byte order of the SLM is different from the conventional sequential array storage. The interleaved hologram can be then written into the chosen positions of bit-planes in SLM's flash, and a function (R11_RpcSysReloadRepertoireImageSubset) newly developed by the manufacturer of the SLM (Forth Dimension Displays) is used to update the image storage at the positions.

4.5.2. Hardware triggering for PSF measurements

In imaging the diffracted order, the SLM and the camera can both be operated from Python code. In order to improve the imaging speed and thus the phase recovery iterations, the camera and the SLM are hardware triggered to synchronise the hologram display and the exposure of the camera.

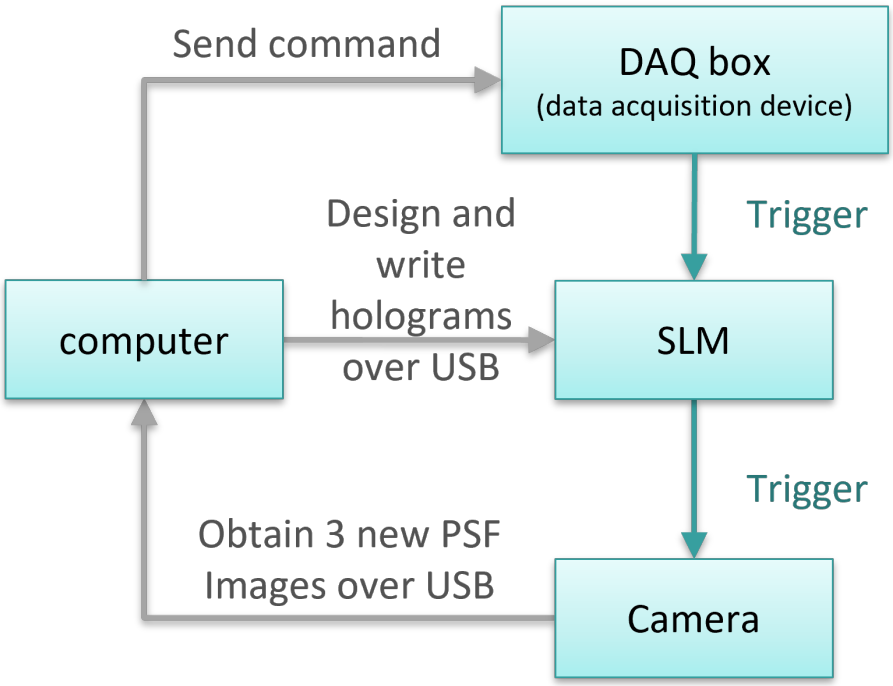


Figure 4-22. Flowchat of the hardware triggering in capturing diffracted orders. All devices are controlled by the Python code by using the corresponding packages. Green arrows denote hardware signals between devices.

As shown in Figure 4-22, three holograms are calculated and then written into the SLM's flash buffer over a USB connection. An acquisition is initiated by a command from the computer to the data acquisition device (DAQ box, NI USB6351) that then sends a trigger to the SLM to start the required running order. Once each required SLM patterns is displayed the SLM drive circuit triggers the camera exposure to acquire the three PSF measurements.

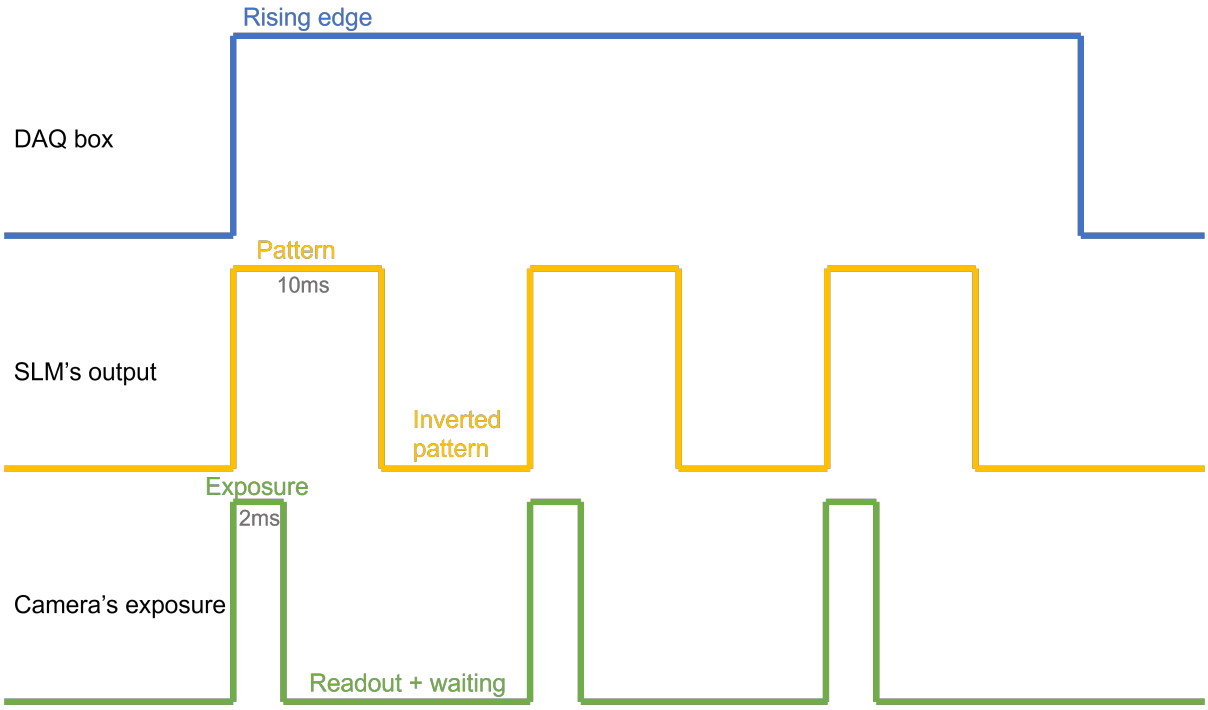


Figure 4-23. Hardware triggering signal of devices. The start of each hologram display is synchronised with the exposure of the camera

The synchronisation between triggering signal is shown in Figure 4-23. The pin 74 (PFI 1/ P1.1) and pin 92 (D GND) of the DAQ box is respectively connected to the port 3 (SPI_1 +) and the port 4 (SPI_1 -) of the SLM to generate a rising edge (the blue signal) and then trigger the SLM to display the three holograms successively, starting from the 0^{th} frame we set. The output of the SLM (the yellow square wave) can be obtained from the port 16 (GND_ISO) and port 17 (LED_EN), corresponding to the three holograms. Due to requirement to DC balance in the SLM, each pattern displayed on the SLM is shown in positive and inverted contrast so that the net voltage seen on the liquid crystal is zero to avoid electrochemical breakdown. The period for displaying each positive and a negative hologram is set by the sequence here as 10ms. The SLM's output signal is connected to the TRIGGER IN connector of the camera's breakout board to trigger the camera's acquisition. The rising edges of the SLM's signal trigger the exposure of the camera (the green signal). The exposure time is set as 2ms, and the camera's readout time for an image with 256×256 pixels is measured as about 7.6ms. Acquiring the PSF images in this way in quick succession is essential so that if there is any extra aberration present that is changing in the system, for instance from air currents, then that aberration changes only minimally between the three PSF measurements. Theoretically, the duration of displaying pattern pairs can be even shorter (as long as the waiting time is long enough to offset the camera's rolling shutter), and the speed of PSF measurements can be further improved.

4.5.3. Measuring and recovering the SLM's phase

In practical measurements, the camera is aligned to locate one light order diffracted from the hologram at the centre of the field of view of the acquired region of interest. After each capture, the image is returned in python as a numpy array. To avoid the impact of offset and noise in the acquired image, a background level set at the 50th percentile p_{50} of all pixel values is subtracted from each image I_i, and only values are above the 95th percentile p_{95} are kept with all the others set to zero. The measured PSF used in the G-S loop is thus calculated as:

$$(I_i - p_{50}(I_i)) \times (I_i > p_{95}(I_i))$$

. As shown in Figure 4-24, the uncorrected PSFs with the three applied biases of $4 \times (B_{w_{02}} + B_{w_{20}})$, $8 \times B_{w_{22}}$ and $-4 \times (B_{w_{02}} + B_{w_{20}})$ are imaged by the camera.

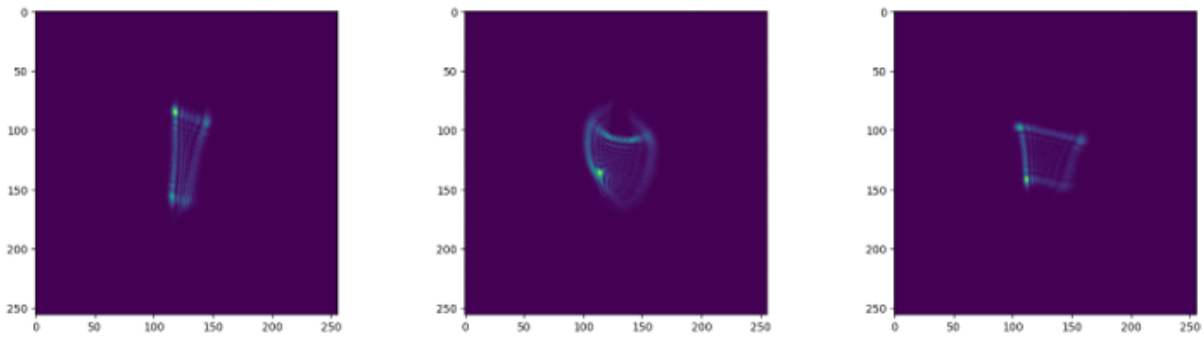


Figure 4-24. The diffraction orders with applied three biases.

Following the initial PSF acquisitions, the phase of the SLM is retrieved by the G-S phase recovery method with the Chebyshev fitting to the result as demonstrated in Figure 4-25. The criterion is that the RMS must be less than 0.15 radians in two consecutive measurements. The three sub-images in the first row are final PSFs obtained after the final correction iteration, and they are similar to the PSFs showed in Figure 4-5, which only include introduced biases. The interleaved hologram sent to the SLM to diffract the first PSF is displayed in (l). As demonstrated in Figure 4-25 (d), the calculated PSF with the first bias (defocus) and is consistent with the measured one (a). The pupil phase and intensity estimated in the last correction iteration are displayed in (e) and (f), respectively. The highest degree of Chebyshev polynomials n_c is set as six, and the matrix of Chebyshev mode values S_0 for the current correction step is reshaped and shown in (g). Most modes contribute a low amplitude in this final correction step. The phase of the piston value that is on the top left (0,0) is deliberately set to zero as it has no effect on the aberration correction. As displayed in (h), the calculated

pupil phase in (e) is unwrapped, and the RMS is calculated as 0.412 radians. The rms of the phase before and after the unwrapping is the same because the retrieved phase in the last correction iteration is small. As demonstrated in (i), Chebyshev polynomials are used to fit the phase, resulting in a significant reduction of the phase RMS to 0.063 radians, which is notably smaller than the diffraction-limited value of 0.469 radians. In the calculation of this rms value the effects caused by the piston at (0,0) and two tilts at (0,1) and (1,0) in the Chebyshev arrays (see $B_{W_{00}}$, $B_{w_{01}}$ and $B_{w_{10}}$ in Figure 4-14) are removed because they can be seriously affected by both the air disturbance and how accurately the PSF is located at the centre of the FOV. Again, these do not contribute appreciably to aberration correction. The rms values of the fitted phase in four correction iterations are plotted in (j), decreasing from about 6.5 radians to 0.063 radians, and the final accumulated phase is demonstrated in (k). The whole algorithm can finish in 20s. While the results remain consistent for the field's phase and essential features (e.g. the size of objective), discrepancies arise primarily from PSF image noise [83] and alignment errors.

The speed of the phase recovery algorithm is mainly limited by the USB transfer speed of the SLM and subsequent transfer from flash storage when updating the three bit-planes. Each correction iteration takes about 4.11s, or about 4.50s when including graphic updates to show intermediate progress results on the fly.

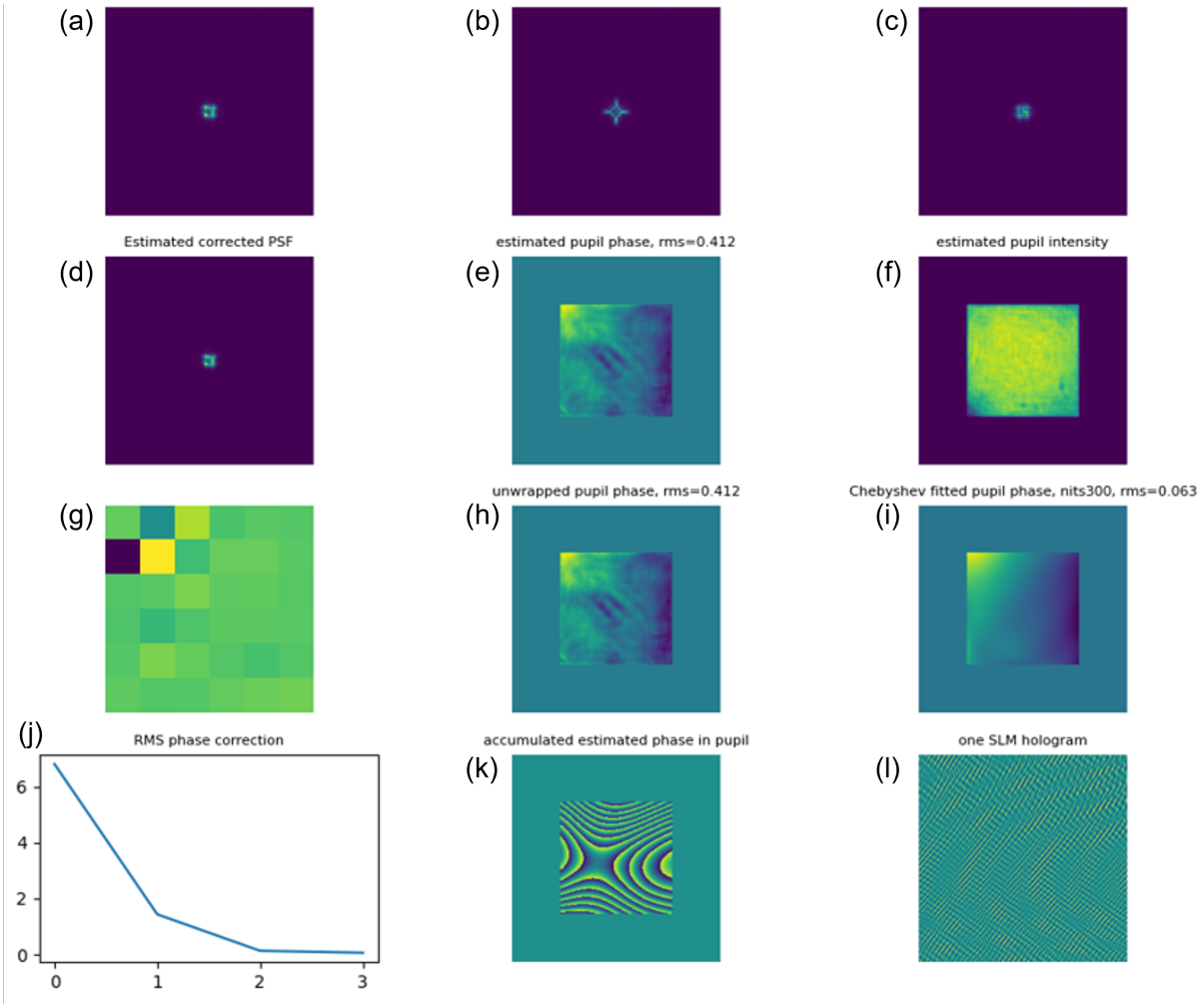


Figure 4-25. Measuring and recovering the phase of the SLM's surface with the simple layout. (a) – (c) Three PSFs with introduced biases. (d) The calculated PSF with the 0^{th} bias. (e) The calculated pupil phase with rms of 0.412 radians. (f) The calculated pupil intensity. (g) The matrix of values of Chebyshev modes in the current step, with the square on the top left showing the piston value. (h) The estimated unwrapped phase with RMS of 0.412 radians. (i) The Chebyshev fitted phase with RMS of 0.063 radians after excluding piston and tilt. (j) Progression of the rms of the fitted pupil phase over the correction process. (k) The total estimated phase accumulation. (l) The hologram written into the flash of the SLM to produce a corrected diffraction order as shown in (a). All results are obtained from the last correction iteration.

As shown in Figure 4-26, compared with the distorted PSF in (a), the diffraction order after the phase recovery (b) is focused to a spot, showing that our heuristic approach is very effective. The corrected diffraction order can completely go through the spatial mask in a SIM system, and the FOV of obtained images can thus be optimised, which will be presented in Chapter 6.

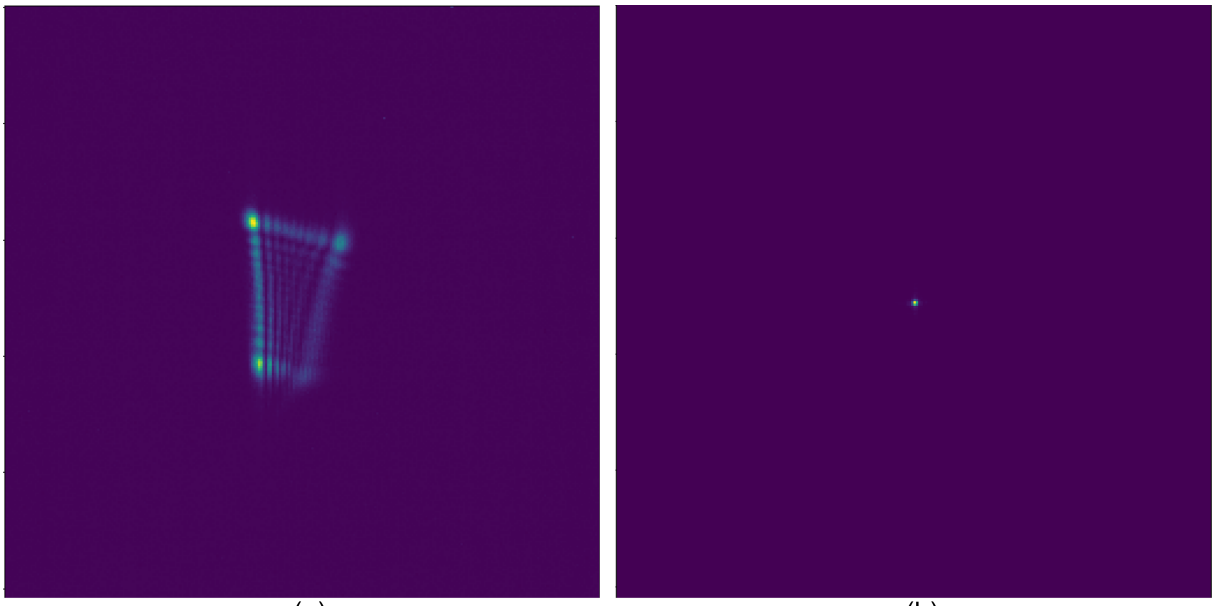


Figure 4-26. Comparison between the initial (a) and the corrected (b) measuremtns of the diffraction order.

The phase correction approach with Zernike polynomials are also implemented and the phase can becorrected and the PSF can thus be recovered. The approach with Chebyshev modes are used in the SIM system since they can completely describe the phase over the square display of the SLM.

4.6. Chapter summary

Extended diffraction orders are noticed in real experiments, which are discovered to be caused by the curved surface of the SLM. This problem demonstrates that the phase the SLM is aberrated, and the desired diffraction orders can thus not pass through pinholes on the mask completely. To address this issue, a phase recovery algorithm is introduced that combines a Gerchberg-Saxton loop based on diffraction measurements and the iterative feedback of partially corrected results from the optics. The phase aberration is too big to be corrected by applying the G-S method alone and there are significant issues with singularities in the retrieved phase. Smooth functions are used to fit the phase to avoid the phase singularities, and the G-S algorithm is applied repeatedly with updated PSF measurements (feedback). The fitted phase can be seen as a partial correction, which is removed from the phase target of the next G-S application. The phase aberration decreases with the iterations of the phase recovery process until it becomes small enough according to a criterion based on the residual rms phase in the current correction. For a square shape of the SLM plane, Chebyshev polynomials are exploited to fit both the amplitude and the phase. Two different expansions: sampled and the

weighted Chebyshev expressions are explained in the chapter. The sampled Chebyshev method defines values at a few positions in the plane and then interpolating them into the entire area, while the weighted one can provide values over all pixels although with weighted values that are empirically adjusted in response to the singularities in the weight function at its edges. In the simulation, there is no noticeable difference in the diffraction orders corrected using two kinds of Chebyshev polynomials, and both corrected PSFs show a diffraction limited result. The RMS is 0.063 radians, which is evidently smaller than 0.469 radians. The scenario with a circular aperture is also simulated, where Zernike polynomials are exploited as the fitting function, and again a heavily distorted PSF is corrected. In practice, the phase recovery approach is realised with a simple optical layout. The diffraction order can be recovered on the fly from the aberrated state to the diffraction limited spot after only 4 iterations in 20 seconds.

Chapter 5. Hologram for HexSIM microscopy and aberration correction

Although there are many approaches that may be used to generate the illumination patterns in a SIM system, using a spatial light modulator (SLM) to display holograms is a very favourable choice, since this method can provide different illumination patterns faster and more accurately compared to using physical gratings, but also makes correcting some optical aberrations possible—an advantage compared with the faster approach of using optical electro-optic modulators. In this chapter we will discuss three approaches (polka-dots approximation, DBBI method and optimisation by binary search) to design, produce and optimise the holograms used in SLM SIM imaging (corresponding to the SIM illumination patterns) in order to find a most advantageous method for hologram generation. We compare the methods in terms of the intensity of diffraction, the generation speed, uniformity and entirety of structured illumination and relative amplitudes and residual phases of frequency carriers. The effect of utilising pinholes with different sizes is also taken into account in the comparison. In the end, we present an approach of directly binarising the beam interference (DBBI method with aberration correction and phase stepping) to produce holograms quickly, without compromising the uniformity of the structured illumination or the accuracy in determining the frequency carriers.

5.1. Using an SLM to generate SIM patterns

In the earliest generation SIM systems [26] [27], diffraction gratings are used to diffract the light source, and only desired diffraction orders are retained, which interfere to project the structured illumination patterns onto the sample plane. However, to reach higher accuracy and speed of applying the (multiple) illumination patterns, some hardware components such as SLMs, digital micromirror devices (DMDs) and electro–optic modulators [84] can be used. In our system, an SLM is used to generate hexagonal illumination patterns, including corrections for aberrations.

5.1.1. SIM systems using an SLM

More than two decades ago, in 2002, Heintzmann, Jovin and Cremer proposed a theoretical SIM system where the structured illumination was generated using an SLM, and demonstrated the simulated results [85]. One year later, Heintzmann extended the theory further, proposing

a system which illuminated the sample with two-dimensional excitation patterns, still using an SLM, and presented the images with improved illumination [86]. In 2008, Hirvonen et al. showed a SIM system which exploited an SLM to generate structured patterns [87], and the resolution improvements were demonstrated by imaging two proteins in Cos-cells. One year later, super-resolution video microscopy was presented by Kner et al. [88], where a ferroelectric liquid crystal on silicon (LCOS) SLM is used to produce illumination patterns. Since then, SLMs have been becoming a favourable tool for the structured illumination generation in SIM microscopy.

In papers published in 2024, SLMs still played a significant role. For example, Temma et al. used a LCOS SLM to generate the illumination at a selected a plane to suppress the negative impact of the out-of-focus light on the resolution improvement [89]. Using a similar model of SLM, Hannebelle et al. achieved an Open-source microscope add-on for SIM, which makes SIM technique accessible for more institutions and labs [90]. Senftleben et al. also exploited a liquid crystal SLM to realise a volumetric multi-focus SIM system [91].

In our system, to be specific, periodic holograms including frequency components are designed and loaded onto the screen of a ferroelectric liquid crystal (FLC SLM) to act as a physical diffraction grating but with a high flexibility and speed. When the liquid crystal molecules are well organised into layers but are titled at a constant angle σ with respect to the layer normal within each layer [92], the phase is known as smectic C. The phase can transform into smectic C* phase if the direction of chiral molecules within each layer slightly changes to form a helical structure over all layers. The angle between the layer normal and the helical axis keeps being σ. If the material thickness is thin enough (approximately equal to the helical pitch) the helix can be inhibited [93], resulting in the material displaying ferroelectric properties [94]. The permanent dipole moment of each molecule lies in the direction either at $+\sigma$ or $-\sigma$ relative to the layer normal. Ideally each pixel can be regarded as a half-wave plate that switches through between angles of $\pm\sigma$. In this scenario, when linearly polarised incident light is incident on the SLM and aligned to bisect the two switched state axes, then there will be a $\pm 2\sigma$ angle between the azimuths of the incident and output polarisations. Therefore, if a polariser is used to block one of these states, the SLM functions as a binary amplitude modulator for the linearly polarised light. However, if the output polariser is placed orthogonal to the input polariser then both states will produce the same output intensity but with a relative phase shift of π radians [95]. Even if the output polariser is removed the output polarisation can be considered as a combination of an un-modulated component, and a orthogonally polarised modulated component with a relative phase shift of π radians between the two states. For instance, if the

input polarisation is left hand circularly polarised then the modulated output polarisation is right hand circularly polarised. It is this relative phase shift that is used to generate diffracted beams by writing hologram patterns onto the SLM.

In general, each pixel on the SLM's screen can be switched to be either 'on' or 'off' to change the polarisation state of the incident light. The commercial model (Forth dimension displays, R11, 2K) we use can provide a switching speed down to 100 μs, displaying 10000 1-bit 2K hologram frames. The high rate makes the display of different holograms in quick succession and thereby aberration correction on-the-fly (which will be introduced in this chapter) accessible. Holograms are displayed on the SLM's screen to diffract the illumination light into diffracted orders. A lens placed behind the SLM performs the 2-D Fourier transform of the hologram and focuses it on the spatial pinhole mask, which blocks unwanted orders to allow only three beams go through the system. Similarly, the Fourier transform of the interfered beams is projected on the sample plane as the structured illumination. In other word, both the hologram and the illumination pattern are at the Fourier plane relative to the spatial mask.

5.1.2. 3-beam HexSIM illumination

In 2014 Schropp and Uhl demonstrated SIM with two-dimensional illumination patterns such as chequerboard and hexagonal patterns [33]. The illumination in our system is structured as hexagonal, in contrast with the traditional illumination of sinusoidal stripes. The seven holograms required to obtain one image is less than the conventional approach, and the shift in holograms is only along one direction since there is no rotation required.

However, it is important to consider the frequency components included in the hexagonal hologram, which are different from the case of two frequencies in a conventional sinusoidal hologram. In a hexagonal hologram [33], there are three frequency components, denoting the three diffraction directions, and this is the reason why there is no need to rotate hexagonal holograms to extend the resolution isotopically. The electromagnetic field E of a beam is given as

$$E = E_0 \times e^{i(k_x \times x + k_y \times y + p)} \quad \text{5-1}$$

, where E_0 is the initial amplitude, p is the phase of the field, k_x, k_y and x, y are the wavevectors and coordinates in x and y direction, respectively. The hexagonal illumination can be generated by three beams, which can be mathematically described using equation 5-1. The total field E_{total} of 3 beams

$$E_{total} = E_0 \left[e^{i(k_{x1} \times x + k_{y1} \times y + p_1)} + e^{i(k_{x2} \times x + k_{y2} \times y + p_2)} + e^{i(k_{x3} \times x + k_{y3} \times y + p_3)} \right] \quad 5\text{-}2$$

is obtained by adding all the field. The intensity of the total field I_{total} is given by the product of the field and its conjugate ${E_0}^*$, as

$$\begin{aligned} I_{total} &= E_0 \times {E_0}^* \\ &= {E_0}^2 \big[2cos\big(k_{x1}x - k_{x2}x + k_{y1}y - k_{y2}y + p_1 - p_2\big) \\ &+ 2cos\big(k_{x1}x - k_{x3}x + k_{y1}y - k_{y3}y + p_1 - p_3\big) \\ &+ 2cos\big(k_{x2}x - k_{x3}x + k_{y2}y - k_{y3}y + p_2 - p_3\big) + 3 \big] \end{aligned} \quad 5\text{-}3$$

. The three field components generate three intensity components, and their phases $p_1 - p_2$, $p_1 - p_3$ and $p_2 - p_3$ in equation 5-3 can be represented as follows:

$$\begin{aligned} p_1' &= p_1 - p_3 \\ p_2' &= p_2 - p_3 \end{aligned} \quad 5\text{-}4$$

because only the relative phases are considered. The phases in the intensity can thus be described by only two symbols.

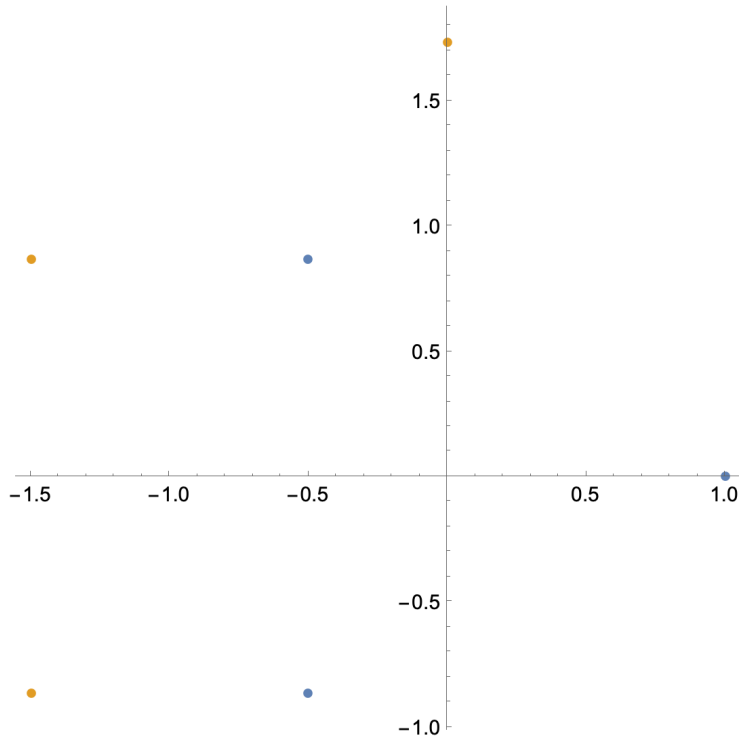


Figure 5-1. Frequencies in the Fourier domain. Three field components (blue) and three intensity components (orange). Conjugate components are omitted from the image.

As shown as blue dots in Figure 5-1, when the three field frequencies (k_{x1}, k_{y1}), (k_{x2}, k_{y2}) and (k_{x3}, k_{y3}) for generating the hexagonal pattern are given as $(-\frac{1}{2}, \frac{\sqrt{3}}{2})$, $(-\frac{1}{2}, -\frac{\sqrt{3}}{2})$ and $(1,0)$, the corresponding intensity I can be obtained as by substituting the three field vectors into equation 5-3, and the resultant intensity is

$$I = \frac{3}{2} + cos[p_1' - p_2' + \sqrt{3}y] + cos[p_2' - \frac{3x}{2} - \frac{\sqrt{3}y}{2}] + cos[p_1' - \frac{3x}{2} + \frac{\sqrt{3}y}{2}] \quad 5\text{-}5$$

. The three intensity frequencies $(0, \sqrt{3})$, $(-\frac{3}{2}, -\frac{\sqrt{3}}{2})$ and $(-\frac{3}{2}, \frac{\sqrt{3}}{2})$ can be extracted from equation **5-5**, which are shown as orange dots in Figure 5-1.

5.2. Generating holograms for 3-beam HexSIM illumination patterns

Traditionally, a SIM hologram is designed as a periodic pattern depending on the frequencies needed for reconstructing a super-resolved image. However, these kinds of holograms only provide performance if the optical system is aberration-free. To compare different methods of generating holograms for the real 3-beam HexSIM system, three methods are introduced and discussed in this section.

5.2.1. Polka-dot approximation

In SIM, illumination patterns are periodic, they include a few frequency vectors to be mixed with the frequency components of the Fourier transform of raw images to extend the support of the OTF of the system. Therefore, the hologram uploaded on the SLM can be regarded as an enlarged or reduced illumination pattern projected on the sample plane, depending on the magnification of the optical system. For a SIM system using interfered three beams to generate the hexagonal structured illumination, the frequency pattern is a 2-D image of hexagons consisting of polka dots. According to the positions of frequency dots, the hologram can be generated as an image of polka dots to efficiently diffract the light into multiple beams.

In order to facilitate the alignment between the diffraction orders and the pinhole mask, the Cartesian coordinates x and y are rotated at a definable angle θ. The rotated coordinates x_r and y_r can be calculated as the dot product of a rotation matrix R and the Cartesian coordinates.

$$\begin{pmatrix} x_r \\ y_r \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad 5\text{-}6$$

The period in the horizontal and vertical direction is set as p and $\frac{\sqrt{3}p}{2}$, respectively [33]. p is the basic period of the pattern, denoting the minimum distance between two polka dots. With the coordinates, the number positions of centres of the polka along the y axis are chosen as integer multiples of $\frac{\sqrt{3}p}{2}$, starting from 1 and keeping in a column:

$$y_i = \left\lfloor \frac{y_r}{\frac{\sqrt{3}p}{2}} + 0.5 \right\rfloor \quad 5\text{-}7$$

, where $\lfloor a \rfloor$ takes the modulo of a and i is the number of positions. The corresponding x centres are determined as

$$x_i = \left\lfloor \frac{x_r}{p} - \left(\frac{y_i \% 2}{2} + 0.5 \right) \right\rfloor + \frac{y_i \% 2}{2} \quad 5\text{-}8$$

, where i is the position number and $\%$ is the remainder operator. The difference in the number positions between odd and even y_i is 0.5. Multiplying the number position with the periods in two axes, respectively, the actual polka dots positions x_0 and y_0 can be obtained as

$$y_0 = y_i \times \frac{\sqrt{3}p}{2}, \qquad x_0 = x_i \times p \quad 5\text{-}9$$

As shown in Figure 4-21 (a), the basic periods of dots centres in x and y are showed with the horizontal and the vertical arrows, and the horizontal shift between two adjacent rows is $\frac{p}{2}$.

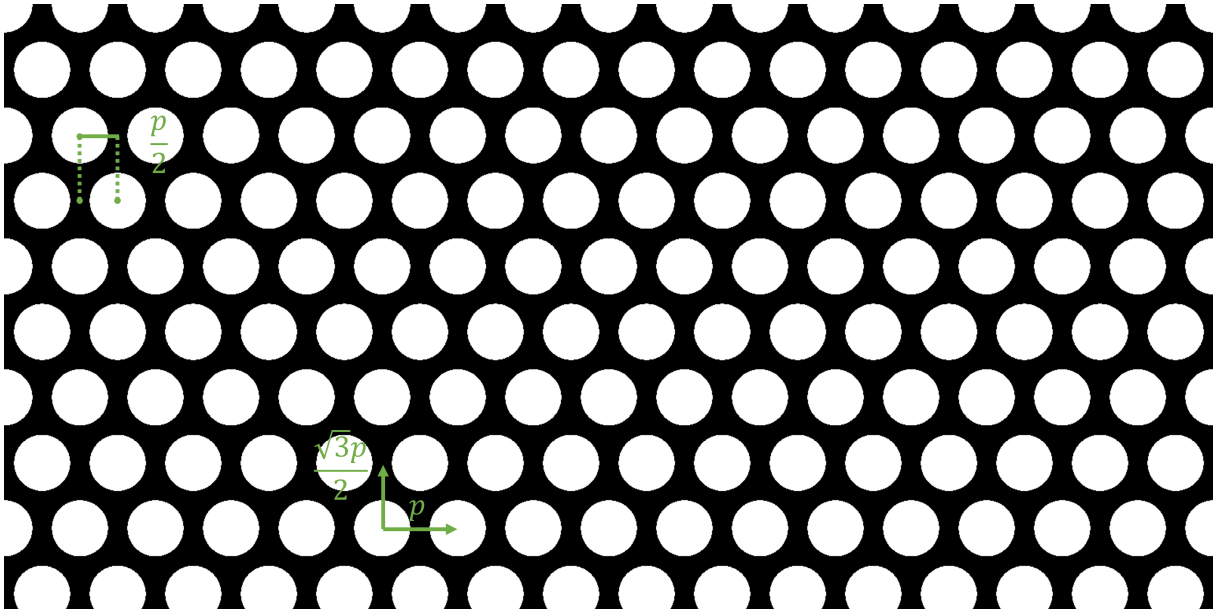


Figure 5-2. Polka dots. Periods of the pattern in x and y directions are annotated.

After determining the positions of polka dots, the radius of each polka dot r can be written as

$$r = \sqrt{(x_r - x_0)^2 + (y_r - y_0)^2} \quad 5\text{-}10$$

. However, the radius r_0 that makes the maximum diffraction of the hologram still requires to be determined. Since the pattern of diffraction beams is the Fourier transform of the hologram, the zeroth order can be seen as the DC term. For a binary hologram, the DC term is the integral of phases of the SLM as

$$\int e^{i \times phase} \quad 5\text{-}11$$

, where i is the imaginary unit. The binary phase in the SLM plane is either 0 or π, and $e^{i \times phase}$ can thus be -1 or 1, respectively. To diminish the DC value for the higher energy for other diffraction orders, the numbers of pixels with 0 and π in the hologram should be the same. Namely, as shown in Figure 5-3 (a), the black and white areas in a basic triangle required to be the same. To this end, the relationship between the horizontal period p and the radius r_0 is calculated by taking the summed areas of three circular sectors as the half of the area of the basic triangle as

$$3 \times \left(\frac{\pi r_0^2}{6}\right) = \frac{1}{2} \times \frac{1}{2} \times p \times \frac{\sqrt{3}p}{2} \quad 5\text{-}12$$

, so $r_0 = \frac{1}{2}\sqrt{\frac{\sqrt{3}}{\pi}}p$. Given the definition of the radius in 5-10, the polka dots in a hologram can be produced with r_0 by only taking pixels inside polka dots ($r < r_0$) as 1 and keeping the remains as 0 or vice versa. In this way, polka dots can be drawn onto the ‘digital paper’ pixel by pixel.

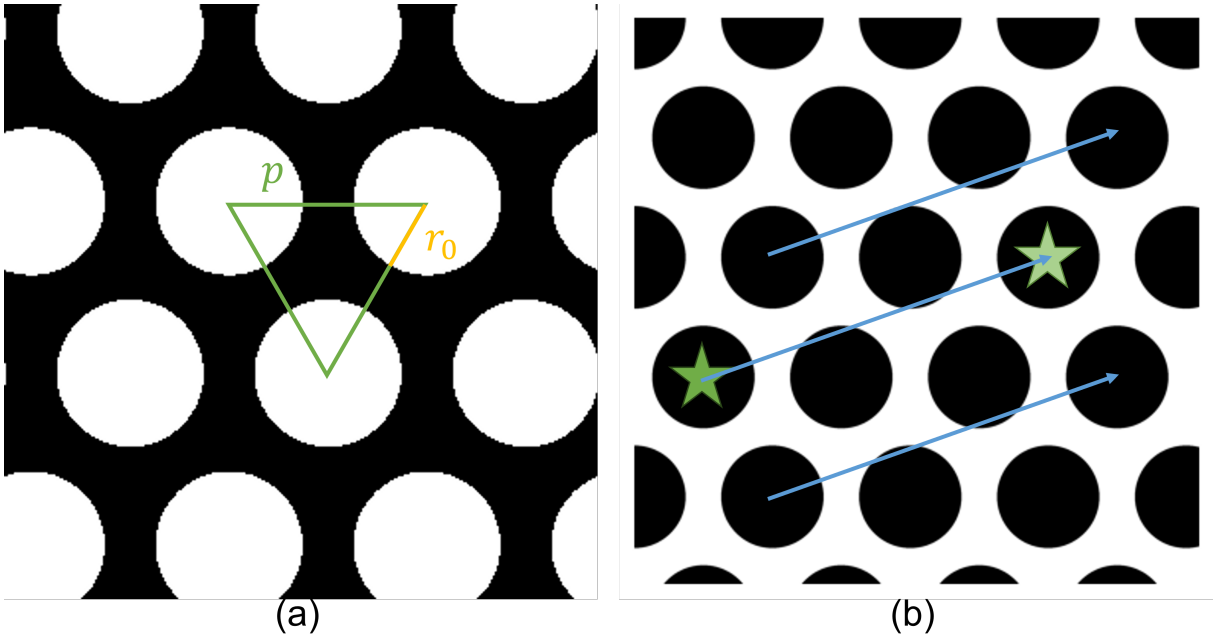


Figure 5-3. Parameters of dots pattern. (a) The relationship between the pitch and the radius of a polka dot. p and r_0 is annotated in green and yellow, respectively. The image is cropped from Figure 4-21 and is zoomed in. (b) The shift in seven patterns is displayed by blue arrows. The dark and light green star denotes the original and the final position of a polka dot, respectively. The positive and negative areas are opposite to (a), but the pattern is still binary.

As the number of illumination patterns in a HexSIM system to reconstruct a super-resolved image is seven, the shift in each pattern thus requires definition. The physical shift in the patterns is due to the change in the phase of beams which interfere to generate illumination patterns. The phase shift of each hologram as $h \times \frac{2\pi}{7}$, where h is the hologram number. Correspondingly, the spatial shift of each hologram is the seventh of the movement from frame 0 to frame 6. As demonstrated in Figure 5-3 (b), the blue arrows show the movement of polka dots in seven patterns, moving $\frac{5}{2}$ period and 1 period in x and y direction, respectively. The dark green star denotes the position of the centre of a polka dot in frame 0, while the light green one annotates the location of the polka dot in frame 7 that is the starting frame of the next stack. This hologram shifting is shown in video 'hologram_shift' [96] (Structured Illumination Microscopy - Thesis supplementary videos (figshare.com)). Therefore, each spatial shift can be proportionally decomposed along the two directions as $\begin{pmatrix} \frac{-\frac{5p}{2} \times h \times \frac{2\pi}{7}}{2\pi} \\ \frac{\frac{\sqrt{3}p}{2} \times h \times \frac{2\pi}{7}}{2\pi} \end{pmatrix}$. The two coordinates with the shifts for one hologram (x_h and y_h) can thus be written as

$$\begin{pmatrix} x_h \\ y_h \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -\frac{5p}{2} \times \frac{h}{7} \\ \frac{\sqrt{3}p}{2} \times \frac{h}{7} \end{pmatrix}, \qquad h = 0,1,2,\cdots 6 \tag{5-13}$$

. The coordinates in equation 5-6 are hence substituted with those in 5-13 to generate a stack of seven holograms. The radius of each polka dot r can be written as

$$r = \sqrt{(x_h - x_0)^2 + (y_h - y_0)^2} \tag{5-14}$$

, and a binary hologram can be generated by setting the area where $r > r_0$ as 1 and the remaining as 0, respectively, or vice versa. For binary holograms, the white and black areas can be swapped without affecting the correspondent diffraction. Figure 5-4 shows the stack of zoomed-in cropped seven holograms with the spatial shift.

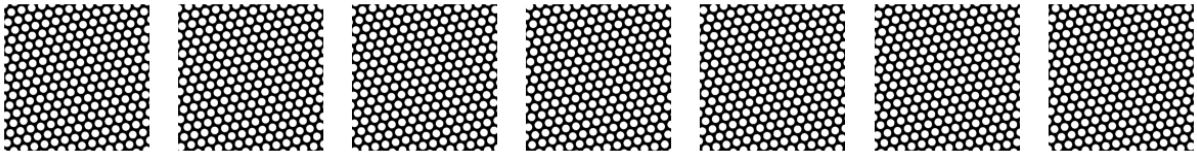


Figure 5-4. Zoomed-in cropped seven holograms produced by the polka dots approximation. The holograms are rotated with $\frac{\pi}{10}$ rad to show the sptial movement more clearly.

However, producing holograms on a pixellated device is different from drawing polka dots patterns. In practice, the periods of the hexagonal pattern depend on the optical system and this has a large impact on the shape of polka dots in a hologram. Hence, it is important to determine the scaling between the SLM plane and its Fourier plane. As shown in a simple case in Figure 5-5, given the angle between the light and the optical axis (z) as θ, the height of the diffraction order in the mask plane can be calculated as

$$h = f \tan\theta \approx f \sin\theta = \frac{f\lambda}{d} \tag{5-15}$$

, where θ can be retrieved from the period of the hologram d and the wavelength of light λ, f is the focal length of the lens performing the Fourier transform.

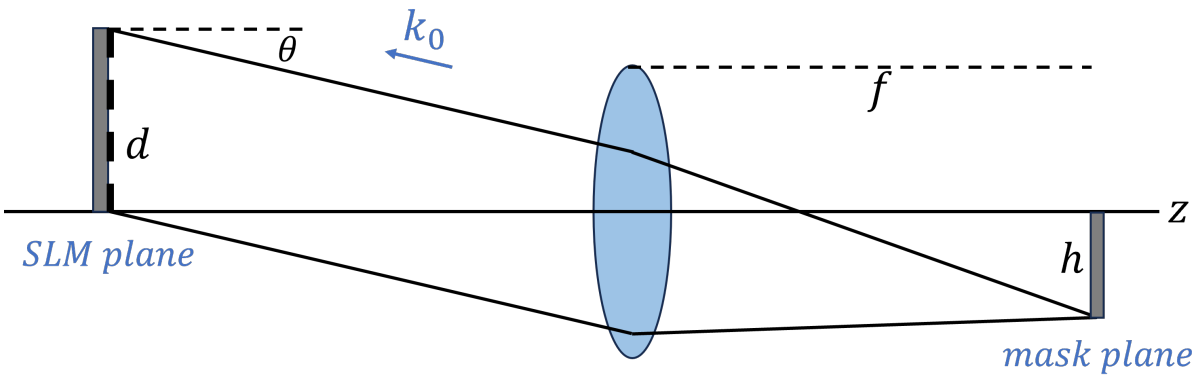


Figure 5-5. Schematic of a lens performing Fourier transform. A hologram loaded on the SLM can be seen as a grating with the period as d.

Compared to the simple case, the scaling between the plane at which the spatial mask locates and the SLM plane is more complex for a hologram with a hexagonal pattern since its Fourier transform result is of three orientations. As shown in Figure 5-6 (a), the light incident on the hologram would diffract the laser light along three directions. Hence, there are three pairs of beams in its Fourier transform, as demonstrated in (b), and the colours of beams refer to the diffraction directions in (a). The conjugates of beams are due to the binary hologram. The grey disc shows the spatial filter in a HexSIM system, where only three beams can go through.

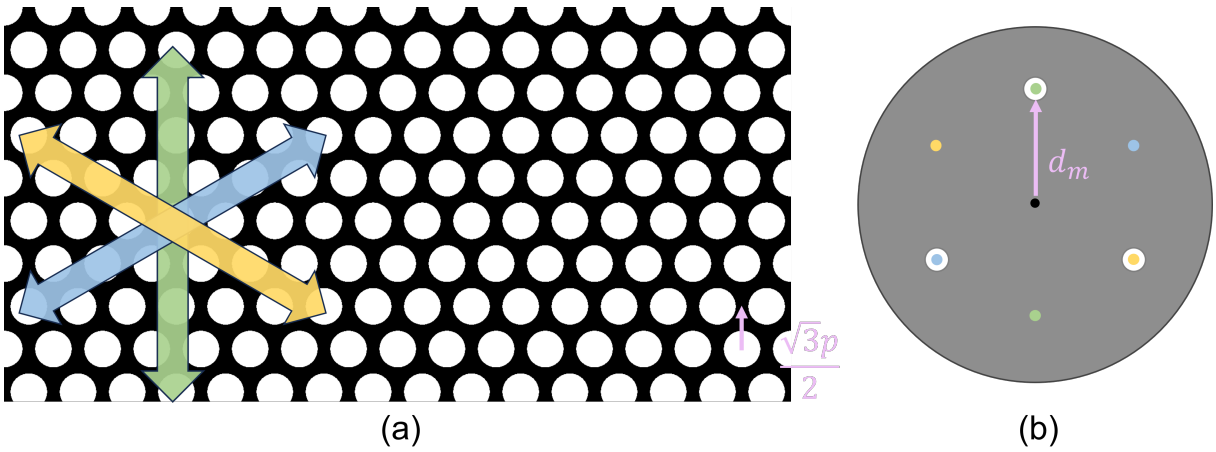


Figure 5-6. (a) A hologram diffracting the light into three directions denoted in different colours. (b) The schematic of the corresponding diffracted orders in the mask (the grey disc) plane. Only zeroth and the first orders are shown. The size of the mask, pinholes and the distance between beams in the image are arbitrarily chosen here to illustrate the layout. The pink arrow in (a) and (b) annotates the distance in the SLM plane and the mask plane, respectively.

Since the radius of the circle including three beams d_m is known in practice, according to equation 5-15, the counterpart $\frac{\sqrt{3}p}{2}$ in the hologram can be calculated as

$$\frac{\sqrt{3}p}{2} = \frac{f\lambda}{d_m d_{SLM}} \quad 5\text{-}16$$

, where d_m is the distance between the optical axis and the diffraction order in the mask plane, and d_{SLM} is the pixel size of the SLM. The pitch of polka dots in pixels in two directions can hence be calculated.

Actual holograms with a calculated pitch don't display obvious polka dots as shown in Figure 5-4 since the number of pixels is finite. The smaller the pitch of polka dots, the less similarity of the hologram with a polka dot pattern. As demonstrated in Figure 5-7 (a) and (b), when the pitch is set as 10 and 8.229 pixels, the dots don't keep the round shape, and the shape is distorted more as the pitch value decreases to 6 pixels in (c). A cropped hologram calculated with parameters in an actual optical set-up is displayed in (b), with λ, f,d_m and d as 561 nm, 250 mm, 2.4 mm and 8.2 μm, respectively.

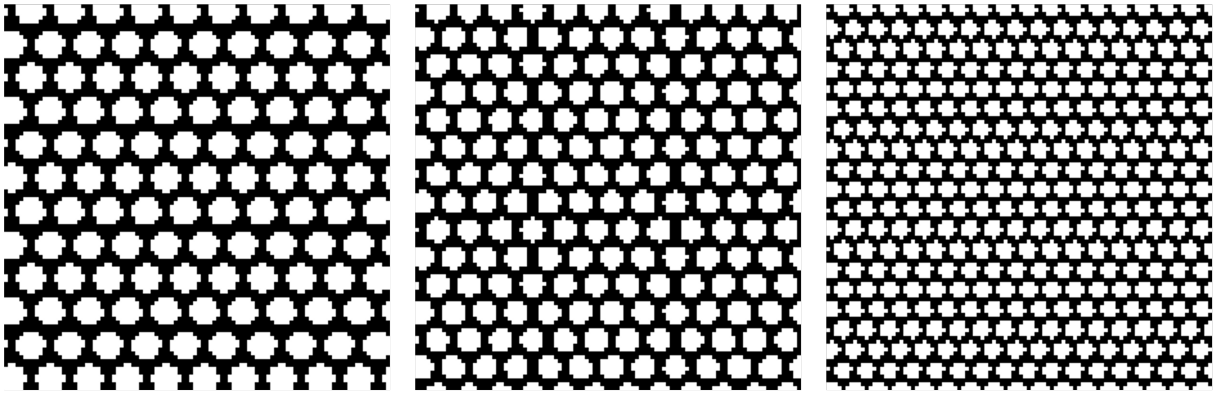


Figure 5-7. Zoomed-in cropped holograms with different pitch p in pixel. (a) p = 10. (b) $p = 8.229$. This is a value calculated with an actual optical layout. $\lambda = 561\ nm$, $f = 250\ mm$, $d_m = 2.4\ mm$ and $d = 8.2\ \mu m$ are substituted in equation 5-16 to calculate the p. (c) p = 6.

In a SIM system, drawing holograms by this method can diffract the light to multiple beams to go through the spatial mask and guarantee the weakest zeroth order. In addition, both the distance between beams and the angle of the beams can be customised according to the specific settings in a SIM layout. However, there is a lack of control of the constant phase of a desired beam among the three in this method.

5.2.2. Direct binarisation of beam interference (DBBI)

In order to generate holograms in a more flexible and bespoke way, we developed a more generic method that is abbreviated as DBBI. Instead of directly generating the periodic holograms, binarising the interference between beams is considered as an indirect approach. Due to the three diffraction beams passed by the spatial mask, there are three wave terms in the hologram plane. The interference with the three light waves can be described by the summation of wave terms. The basic wave vector in the optical system is $k_0 = \frac{2\pi}{\lambda}$. Projecting the basic wave vector to the SLM plane, the vector k_r can be calculated as

$$k_r = k_0 \sin\theta \tag{5-17}$$

. Substituting the vector in SLM plane and the spatial distance d_m in the mask plane in equation 5-15 and 5-17, k_r can be written as

$$k_r = \frac{2\pi f}{\lambda d_m} \tag{5-18}$$

, with the unit being m^{-1}. In the hologram, the coordinates in both two directions (x_2 and y_2) are chosen from $\left[-\frac{d_{SLM}\times n_{SLM}}{2}, \frac{d_{SLM}\times n_{SLM}}{2}\right]$, with the number of coordinates as n_{SLM}. n_{SLM} is the number of pixels along horizontal or vertical direction of the SLM. For instance, n_{SLM} is 2048 for a '2K' SLM.

After determinations of some parameters, the holograms can be produced by summing wave terms according to wanted diffraction orders (beams) in the mask plane. As shown in Figure 5-6 (b), there are three orders with the distance to the optical centre as d_m and the rotation angle between orders as $\frac{2\pi}{3}$. For each beam, the wave vector can be decomposed into two orientations as k_{x_b} and k_{y_b}:

$$k_{x_b} = k_r \cos\left(b \times \frac{2\pi}{3} + \theta\right), \qquad k_{y_b} = k_r \sin\left(b \times \frac{2\pi}{3} + \theta\right) \tag{5-19}$$

, where $b = 0,1,2$ is the number of the beam and θ is the rotation of the hologram (or the three beams, which is the same as θ in the polka dots method. The rotations of the k components in the SLM plane corresponds to spatial shifts in the Fourier space, the position of a beam in the mask plane can thus be deduced as the centre position plus an added shift. The spacing k_s in the mask plane (that is also the PSF plane of the SLM plane) can be calculated as

$$k_s = \frac{f\lambda}{n_{SLM} d_{SLM}} \tag{5-20}$$

. Given the spatial distance between the mask centre and a beam d_m, the corresponding distance r_{cen} in pixels can be calculated as

$$r_{cen} = \frac{d_m}{k_s} \quad 5\text{-}21$$

. The position of each beam can be determined as

$$(x_b, y_b) = \left(\frac{n_{SLM}}{2} + r_{cen}\cos\left(b \times \frac{2\pi}{3} + \theta\right), \frac{n_{SLM}}{2} + r_{cen}\sin\left(b \times \frac{2\pi}{3} + \theta\right)\right) \quad 5\text{-}22$$

. The field of the wave can be calculated as an exponential term of the tilted phase as

$$f_b = e^{i\left(k_{x_b}x_b + k_{y_b}y_b + \varphi_b + \varphi_{con}\right)} \quad 5\text{-}23$$

, where φ_b is the phase of each beam and φ_{con} is the constant phase. Considering the phase shift in seven holograms, as explicated in [33], the phase of three beams should be $\varphi_{0h} = 0$, $\varphi_{1h} = h \times \frac{2\pi}{7}$ and $\varphi_{2h} = h \times 3 \times \frac{2\pi}{7}$, where $h = 0,1,2, \ldots 6$. Besides, a phase φ_c can be deduced as the correction explained in Chapter 4. The field of the waves of each hologram f_h can be written as the sum of three wave terms:

$$f_h = \sum_{b=0}^{2} e^{i\left(k_{x_b}x_b + k_{y_b}y_b + \varphi_{bh} - \varphi_c + \varphi_{con}\right)} \quad 5\text{-}24$$

. The hologram can be generated by binarising the field of the interference between three waves, where the area of $real(f) > 0$ is set as 1 and that of $real(f) < 0$ is set as 0, or vice versa, where $real(f)$ is taking the real part of f.

Because of the range of parameters mentioned above (such as the number of beams b, the positions of the beam (x_b, y_b), the control of phases and so on), this method can be used to produce a range of holograms. In Chapter 4, the holograms diffracting the light to two orders are generated using this approach. However, although the phases between beams can be adjusted in this approach, some changes are introduced in the holograms. As displayed in Figure 5-8, compared to holograms in Figure 5-4 with only a change in positions of dots, the shape of dots and the arears of black and white in each hologram are different. As show in the video 'holohgram_shift_with_different_beam_phases'[96] (Structured Illumination Microscopy - Thesis supplementary videos (figshare.com)), the shape of dots changes from circles to triangles in seven holograms, with both the correction phase φ_c and the constant phase φ_{con} being zero.

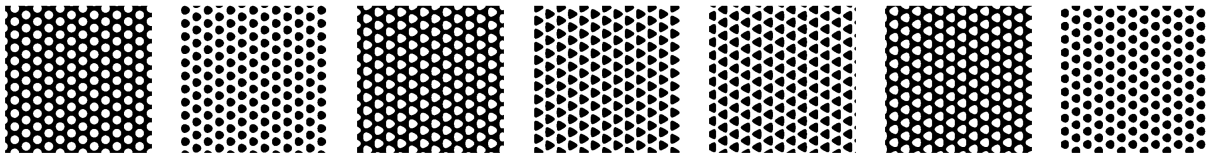


Figure 5-8. Zoomed-in cropped seven holograms produced by the binarisation of the interference between three beams. The bemas phase φ_{bh} changes with the number of hologram while both φ_c and φ_{con} keep as zero.

In addition, the shape of dots changes with different constant phase φ_{con}. As shown in Figure 5-9, while keeping compared to the case with zero constant phase in (a), the dots in (b) are triangles, and the shape of dots in (c) and (d) is round and triangle, respectively. All In fact, the shape of dots in the pattern has an impact on the diffraction intensity of a hologram.

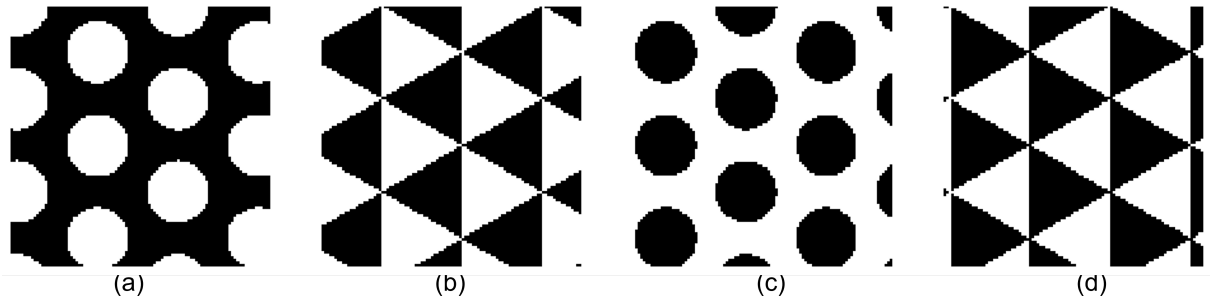


Figure 5-9. Zoomed-in cropped holograms with different constant phase φ_{con}. (a) $\varphi_{in} = 0$. (b) $\varphi_{in} = \frac{\pi}{6}$. (c) $\varphi_{in} = \frac{2\pi}{6}$. (d) $\varphi_{in} = \frac{3\pi}{6}$. The correction phase φ_c for all holograms is zero.

5.2.3. Holograms optimised by binary search

Another alternative approach to hologram generation is to iteratively search for the value of the hologram pixel by pixel with a target image of desired diffraction orders. Methods of directly designing holograms according to intensities of spots were introduced decades ago, such as in [97] and [98]. The approach discussed here is designed based on the direct binary search algorithm in [97], and the implementing code is written by Prof. M.A.A.Neil.

Compared with binarising the interference between the beams explained above, the direct binary search directly generates holograms with defined amplitudes of the wanted diffraction orders, excluding the calculation of the complex values in the mask plane. To begin with, the number, locations and intensities of the desired diffraction orders are defined as the target values, which are consistent with the three beams located at the three pinholes on the mask. After that, the initial hologram is set with random pixel values, and each pixel is either flipped

or keeps its value to minimise the difference between calculated values in the mask plane and the target. The generation of one hologram is shown in the video 'process_of_hologram_generation_by_binarisation_by_direct_search' [96] (Structured Illumination Microscopy - Thesis supplementary videos (figshare.com)), where a hologram is produced by scanning each pixel for the minimum error between the reconstruction beams and the target.

In addition, the approach can be varied for application to many situations. For instance, the target diffraction orders can be of different intensity weights, the potential optical aberrations between the hologram plane and the mask can be corrected, and so on. As presented in Figure 5-10, the level of the holograms is not limited to two (a), where the hologram with four and eight levels are shown in (b) and (c), respectively. In this way, there are no conjugates of the desired diffraction orders, giving rise to a high diffraction efficiency. These holograms with levels higher than two can be exploited with phase-only SLMs that capable of modulating continuous phases, which are essential in fields requiring precise phase control, such as digital holography, adaptive optics, beam shaping and optical trapping and tweezers.

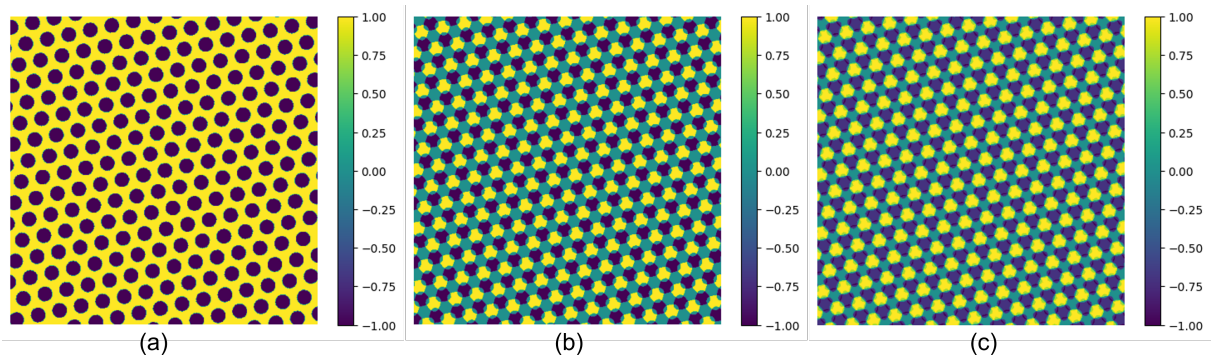


Figure 5-10. The hologram with the phase level as 2 (a), 4 (b) and 8 (c), respectively. All images are obtained by taking the real part of the field retrieved by the binary research.

However, the drawback of the direct binary search is obvious. The speed of this iterative method is low, especially with a large hologram, since the error or the criterion is required to be calculated at each pixel. Therefore, it is also not practical to exploit this algorithm to produce a large number of holograms in-situ to cope with different scenarios such as optical aberrations or adjustments of optics in the system.

5.3. Diffraction efficiency and the pixel balance problem

Given the three methods for the hologram generation, it is worth discussing advantages and drawbacks of each of them. Measuring the diffraction efficiency of a hologram is significant since it will determine the magnitude of the illumination pattern projected on the sample. We start our analysis by examining an ideal SIM system not including any optical aberrations, and then push the discussion towards an aberrated situation.

5.3.1. Aberration-free situations

To measure the diffraction efficiency of a hologram, the intensities of the diffraction orders (PSFs) are calculated, and the balance of positive and negative pixels is taken into account. The diffraction orders and the correspondent diffraction intensity can be simulated.

To the convenience, the holograms are simulated with the size of 500×500 pixel. The PSFs can be written as the Fourier transform of the hologram as

$$psfs = FT\left[w \times e^{i(\pi \times hol)}\right] \quad 5\text{-}25$$

, where w is a Blackman window function used to suppress sidelobes in the Fourier image and hence suppress edge effects in the image of the SLM at the sample., $i = \sqrt{-1}$, and hol is the hologram. Before comparing results from the three methods, it should be noted that the aliasing occurring on the mask plane can be mitigated by rotating the hologram to a little extent. To demonstrate the impact of the rotation on the aliasing, the hologram in the stack generated by the polka dots approximation with and without rotating $\frac{\pi}{10}$ is substituted in the equation 5-25, respectively, and the aberration is set as zero, assuming the optical system is aberration-free.

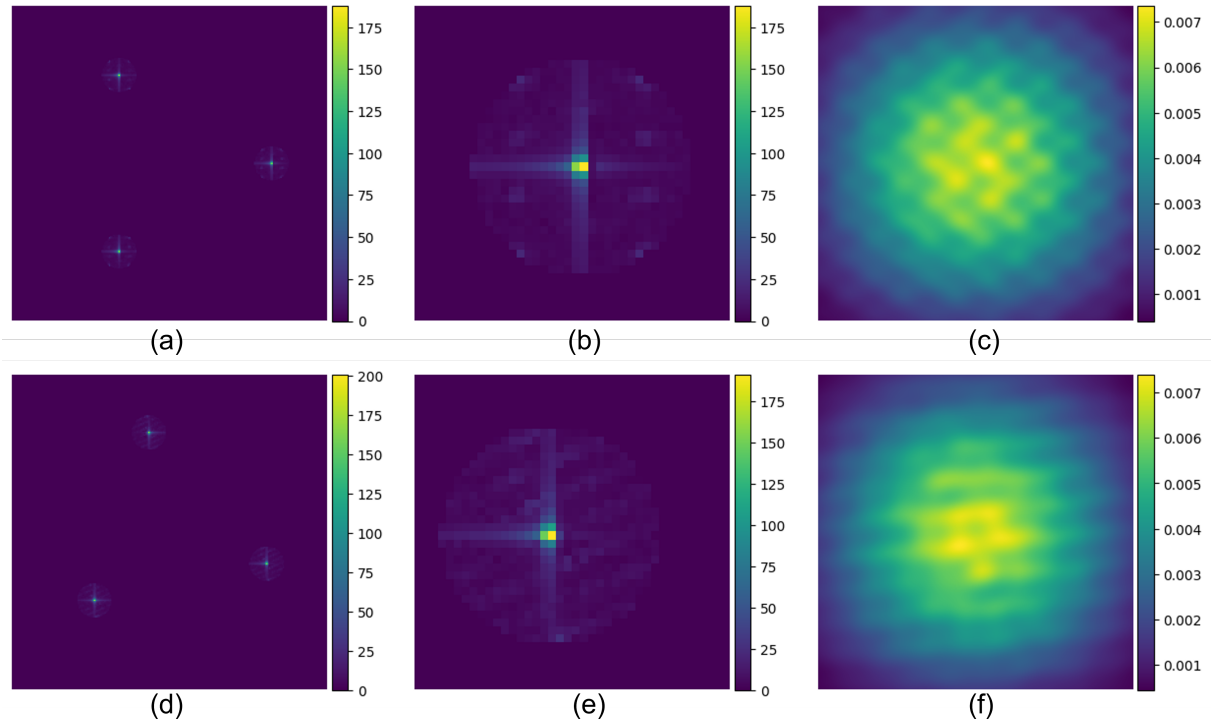


Figure 5-11. Diffraction orders and the correspondent illumination. (a) shows the diffraction patterns passing through the three pinholes. (b) shows the zoomed in pinhole which locates on the right, and (c) displays the illumination from the pinhole. As the hologram is rotated with $\frac{\pi}{10}$, the three orders shown in (d). (e) and (f) show the zoomed in beam and the illumination after the rotation, respectively. All results are calculated with holograms generated by the polka dots approximation.

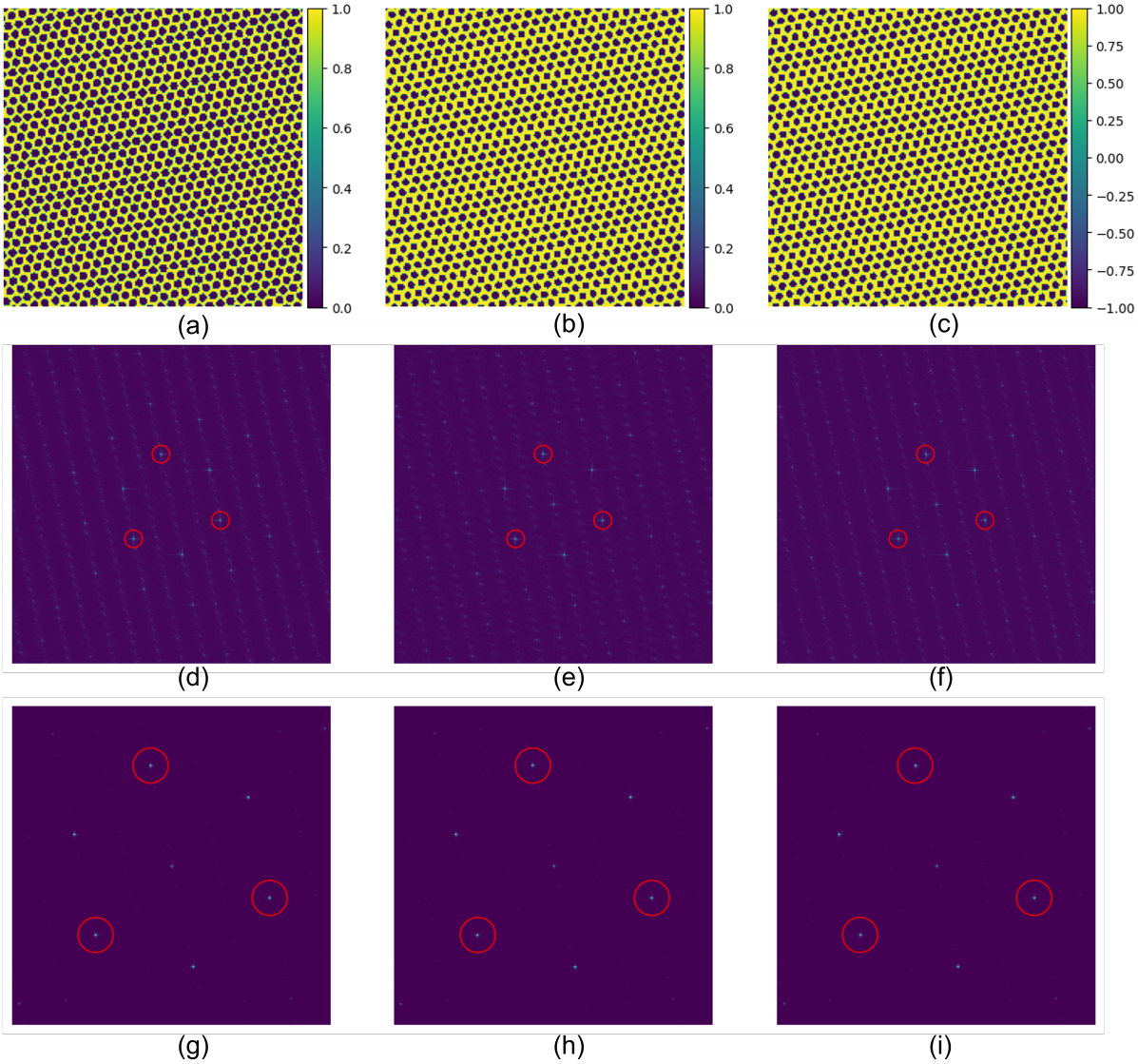


Figure 5-12. Results produced using three hologram generation methods. The hologram is generated by (a) polka dots approximation, (b) DBBI method and (c) the optimisation of the binary search, respectively. (d), (e) and (f) show the square root values of the diffracted field to highlight diffraction orders. The absolute values of the field of diffraction beams are shown and zoomed in further in (g), (h) and (i). All images are cropped and enlarged to show the patterns more clearly. Red circles denote the positions of pinholes on the mask.

The intensities of going through three pinholes before and after the rotation are shown in Figure 5-11 (a) and (d), respectively. As shown in the magnified pinhole in (b) and (e), there are other higher orders present apart from the desired one, resulting from aliasing. When the hologram is not rotated, the unwanted higher orders are aligned in both horizontal and vertical directions, giving rise to the interference displayed in (c). However, after changing the orientation of the hologram, the higher orders in the pinhole are out of alignment (see (e)), and the striped interference pattern in the illumination shown in (f) is reduced. Therefore, introducing a small rotation to the mask installation is a useful strategy in practical experiments.

As shown in the first row in Figure 5-12, three holograms with the rotation of $\frac{\pi}{10}$ are thus generated using the polka dots approximation, DBBI and the optimisation using binary search, respectively. The yellow and the dark purple areas in the holograms generated by the first two methods denote values of 1 and 0, respectively, although the ratio values of two areas in (a) and (b) are different. The highest and lowest value in (c) is π and $-\pi$, respectively, since the field of the hologram plane is retrieved in the binary research optimisation. The square root values of the field of diffraction orders in three cases are displayed in the second row to show the higher orders more obviously. Compared to (e) and (f), the zeroth order in (d) is weak, which is because of the equal areas of black and white pixels. The order located at the centre can be observed in practical scenarios since the size of the screen is limited, although it can hardly be seen in the calculation (d). The absolute value of the field of diffraction beams are enlarged and presented in the third row. The zeroth order cannot be shown in the result from the polka dots approach (g), and the other two results in (h) and (i) perform similarly.

Given the field of PSFs in equation 5-25, the diffraction intensity D of a hologram can be calculated as the summation of the passed light intensity as

$$D = \sum pinholes \times |psfs|^2 \qquad 5\text{-}26$$

, where $pinholes$ is the function of three pinhole windows, and $psfs$ is the field of PSFs. Therefore, the diffraction efficiency of three methods of hologram generation can be calculated and examined, although this definition includes unwanted (parasitic orders) passed by the pinhole. Recalling that the shape of dots in holograms changes with the constant phase φ_{con} in the DBBI method, the diffraction intensity of holograms generated with different constant phases can be studied as well. To explore the change in diffraction intensity with the constant phase φ_{con}, φ_{con} is varied between 0 and π with 102 steps in order to generate 102 holograms. The two holograms produced by the other methods are analysed along with these 102 holograms. As shown in Figure 5-12, the diffraction intensity of the polka dots approximation (the green line) is the lowest, while the value of optimisation by binary research (the red line) is highest, being about 1.088 times the value of polka dots approach. Denoted as the blue line in Figure 5-13, the diffraction brightness obtained using the DBBI method fluctuates with the constant phase φ_{con} with a period of $\frac{\pi}{3}$. The diffraction brightness is the strongest as φ_{in} equals to 0, and the value is the least when $\varphi_{in} = \frac{\pi}{6}$. The change in the diffraction intensity can cause unsteady illumination patterns in SIM, and thus leads to artefacts in reconstructed results. Referring to Figure 5-9, the diffraction efficiency of a hologram is the

maximum when the shape of dots is round, while as the dots are triangles the diffraction is a minimum. The minimum diffraction intensity is about 0.934 times the polka dots result, and the maximum is around 1.090 times.

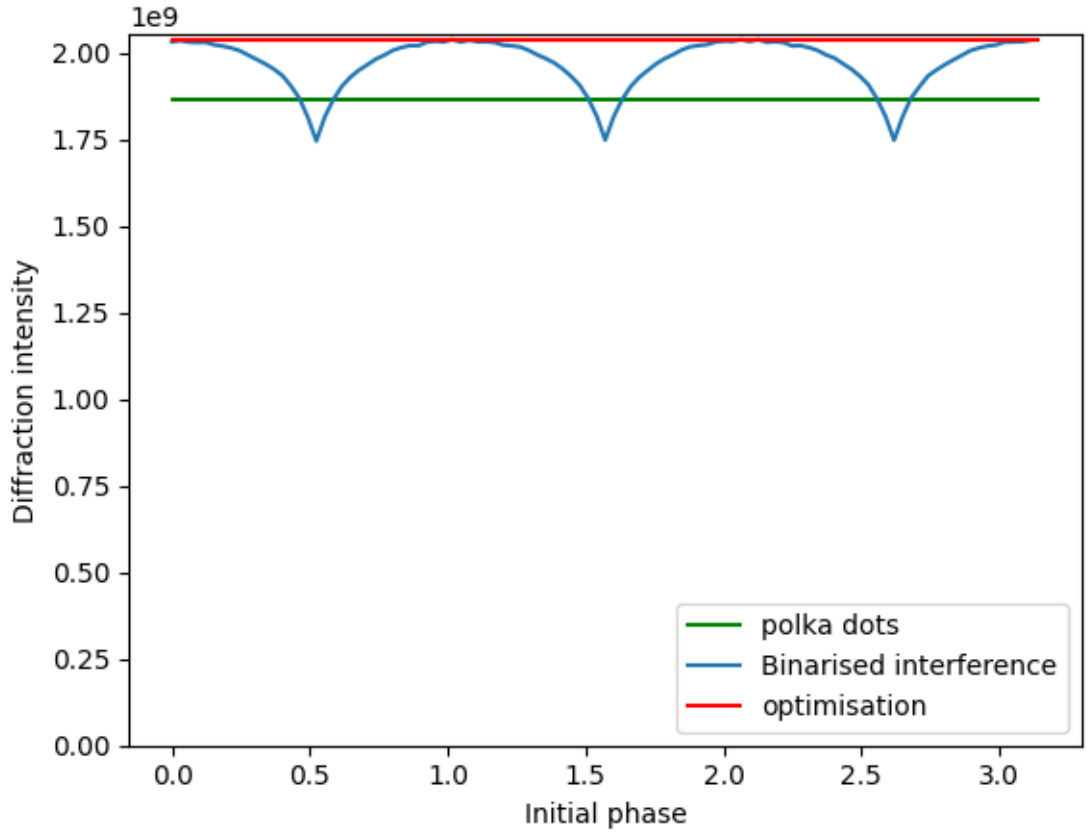


Figure 5-13. Comparison of the diffraction intensity between the polka dots method, DBBI and the binary search optimisation.

In terms of the pixel balance, the non-dots areas (yellow parts in holograms) take 50.015% of the whole area of the hologram in the case of polka dots approximation method because this is deliberately restricted in equation 5-12. Surprisingly, the ratio between the yellow and the dark purple areas in the hologram is (approximately) the same in DBBI with $\varphi_{in} = 0$ and the optimisation approach, with the yellow areas being as 60.149% and 60.137% of the hologram, respectively. When the ratio calculation in equation 5-12 changes so the areas of dots and of non-dots reaches 4:6, the diffraction intensity is around 1.089 times the original value, which is approximately the same as the highest values in the other two methods. Therefore, to achieve the highest diffraction efficiency in the polka dots approximation case, the ratio of the dots area should be 40% rather than the half of the whole area. In a nutshell, the maximum diffraction intensity is obtained in all three methods of generating holograms with this ratio.

As a preliminary conclusion, for a SIM system without optical aberrations, the highest diffraction efficiency can be obtained using each of the three approaches with setting the 4:6 ratio in the polka dots approximation, although we need to be careful about the fluctuation in the diffraction intensity with the constant phase in DBBI.

5.3.2. Cases with aberration and speed of hologram generation

When there are optical aberrations in the SIM system however, the performances of the three holograms generation methods will change. Here, we simulate a scenario where the aberration causes a diffraction order to extend (and hence not completely pass through the pinhole on the mask) before applying the three different methods of hologram generation to assess the diffraction efficiency of each, respectively.

The retrieval and correction of the aberration introduced by the non-flat SLM are explained in 4.1. As shown in Figure 5-14 (a), the aberration originating from an SLM with a non-flat surface is simulated, and the corresponding correction is calculated by the approach in Chapter 4, as displayed in (b). Although the approach is empirical, the difference between the aberrated phase and the correction is small, as presented in (c).

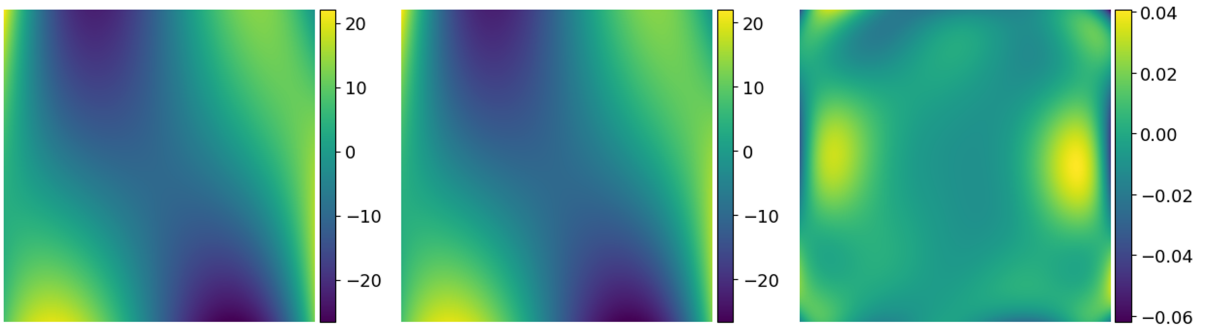


Figure 5-14. Comparison between the simulated aberration and the correction. (a) The aberration in the SIM system. (b) The retrieved correction. (c) The difference between the correction and the aberration. The unit of colour bars is radians.

By using DBBI approach, the aberration can be addressed by substituting the phase of correction φ_c (see equation 5-24) with the correction shown in Figure 5-14 (b), and the extended diffraction orders can be recovered to diffraction limited ones. The holograms generated by the three approaches are displayed in Figure 5-15. As shown in (b), the pattern

is obviously adjusted to include the aberration correction, however, the yellow area comprises 50.12% of the total area. This means that in a system with some aberrations, DBBI method can provide a lower zeroth order than the other two. While the positive and negative areas in the hologram generated by the dots approximation (a) are changed to be 4:6, which is the same as the case of the optimisation method (c).

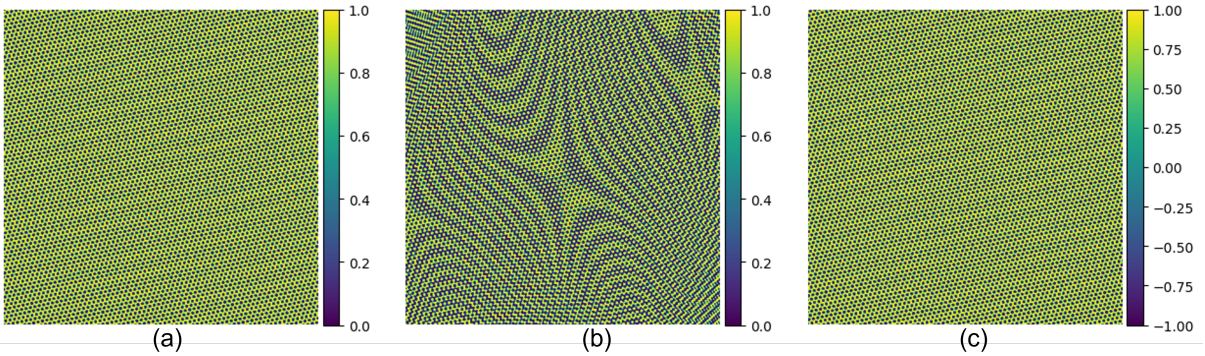


Figure 5-15. (a) The hologram generated by the polka dots approximation with a 4:6 ratio. (b) The hologram generated by DBBI with the phase corection. The yellow area acounts for 50.12% of the whole area. (c) The hologram generated by the optimisation of binary search.

Corresponding to the holograms in Figure 5-15, the diffraction orders are presented in Figure 5-16. The pinhole size is set to permit the main parts of three diffraction ofders to go though, excluding some surrounding extensions. As displayed in the results generated by the 4:6 polka dots in (b) and by the optimisation in (c), the PSFs are distorted and enlarged, which thus cannot pass through the pinholes (denoted by red circles) entirely. Nevertheless, as displayed in (b), the extension of the PSF is corrected by applying the hologram generated by the DBBI method, and the diffraction limited beam can go through the pinhole, completely.

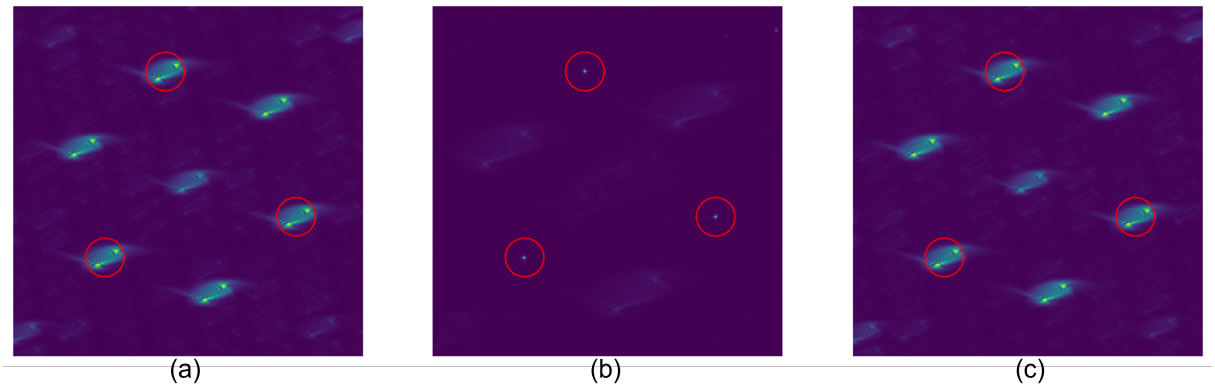


Figure 5-16. The image of the light diffracted by the hologram generated by the 4:6 polka dots method (a), by DBBI including the phase correction (b) and the binary search approach (c). Red circles show locations of the pinholes on the mask.

In terms of the diffraction intensity, as shown in Figure 5-17, the polka dots method with the positive (or negative) area as 50% of the total area performs the lowest brightness when there is the optical aberration in the system. The two methods of polka dots with a ratio of 4:6 and the optimisation of binary presents similar diffraction intensity, which is a little bit higher than the minimum of the vlaue obtianed by the DBBI. With the aberration, the diffraction intensity of DBBI changes with the constant phase as well, however, which is not of the same period as shown in Figure 5-13. The maximum of the diffraction intensity appears in the result obtained by the DBBI, and the fluctuation in the diffraction intensity is mitigated.

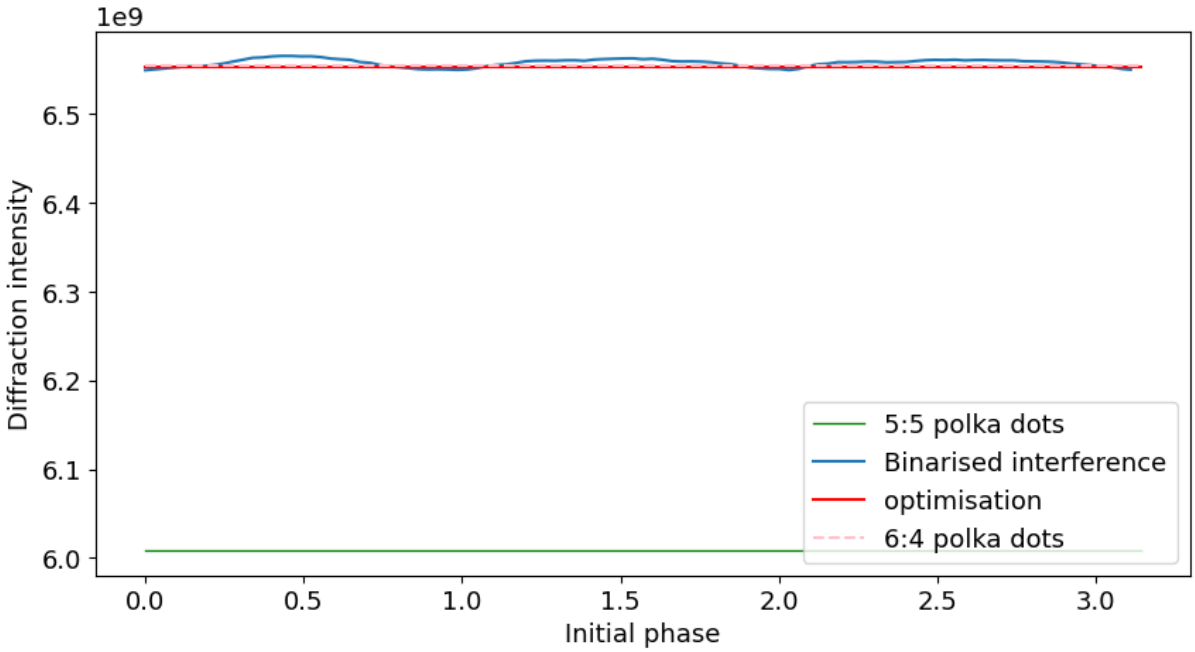


Figure 5-17. Comparison of the diffraction intensity between the 5:5 polka dots method, DBBI, the optimisation and the 4:6 polka dots approach.

Method of hologram generation	5:5 polka dots	DBBI	Optimisation of binary search	4:6 polka dots
Diffraction intensity	1	Max: 1.0928 Min: 1.0901	1.0907	1.0911

Table 5-1. The diffraction intensity of the four scenarios using different ways to generate the hologram. The value of the 5:5 polka dots is set as the benchmark.

As presented in Table 5-1, when the diffraction brightness obtained by the 5:5 polka dots is set as the reference point, the value of the optimisation and the 4:6 polka dots is 1.0907 and 1.0911, respectively. The highest and the lowest value of the binarised interference is 1.0928 and 1.0901, respectively, with the standard deviation as 0.067%. Therefore, the DBBI method can be regarded as the most favourable method to generate the hologram for the high diffraction intensity, although there is a very small oscillation in the intensity.

In addition, it is worthwhile discussing the speed of these methods of generating holograms. To the simplicity, 500 x 500 pixel holograms are generated to test the speed of three methods of designing holograms. As explicated in Table 5-2, when being run on the system (Intell Corel i9-10900K CPU @ 3.70GHz 3.70 GHz), the polka dots approximation is the fastest. However, regarding the superiority of the binarisation of the interference between beams, it is still the most advantageous choice since its speed is still relatively fast. In fact, it is feasible to correct the phase aberration by using the optimisation method because the target function can be tweaked. Nevertheless, this approach is too slower compared to the second method, even though when applying the phase stepping. For example, if we take 100 phase steps in the second approach, 4.7 second is still way much faster than 38.6s. And if the graphical information of the optimisation in the third method keeps updating, the hologram generation is even more time-consuming.

Method	Polka dots	DBBI	Optimisation by the binary search
Speed of generating one 500 × 500 hologram	0.028s	0.047s	Without graphical updates: 38.6s With graphical updates: 61s

Table 5-2. The speed of three methods to generate holograms with a size of 500 x 500 pixel.

By this point we have demonstrated that the method of 4: 6 polka dots is essentially the same as the optimised version. Taking the design time of three approaches into account, the optimisation by binary search is not a favourable choice although it can be extended to

compensate for aberration and to apply phase stepping. Therefore, we will focus on the comparison between the polka-dots approximation and the DBBI method in following sections.

5.4. Uniformity and coverage of illumination at the sample plane

Nonetheless, the highest diffraction intensity of the three beams is not the only criterion to assess the quality of a hologram. To project a superior structured illumination pattern onto the sample in a SIM system, the light diffracted from each pinhole is required to be uniform, and light needs to cover the entire field of view. To compare the three hologram-generation appoaches in term of the uniformity and coverage of the diffraction, the light passing through a single pinhole is firstly studied, the interference between light fields diffracted by three pinholes is then investigated.

To simulate more realistic cases of our HexSIM system, the holograms are generated with the size of 2048 by 2048 pixels to match the real size of the SLM's screen. The window function utalised above is replaced by a Gaussian function to simulate the Gaussian illumination of the illumination laser on the SLM. The three methods are used to generate holograms, with the wavelength of the laser as 488 nm, the focal length of the lens located between the SLM and the spatial mask as 250 mm, the distance between the centre and a pinhole on the mask as 2.4 mm, the radius of a pinhole as $100\,\mu m$, and the pixel size of the SLM as 8.2 μm. Significantly, a typical phase aberration retrieved in our system, as described in Chapter 4, is used in the simulation as the aberrated phase of the non-flat SLM.

For each hologram, the PSF is calculated as the Fourier transform, and the PSF is then multiplied with the pinholes located at the correct positions to simulate the pinhole mask selecting the desired diffractions orders. The parts of the PSF, which can pass through the pinhole, are inverse Fourier tranformed to obtain the light intensity diffracted from the pinhole at the sample plane. Because of the phase recovery in DBBI, as presented in Figure 5-18, the illumination intensity in this case (see(b)) is the most complete and uniform area compared to the situation of the 4:6 polka-dots result (see (a)).The incomplete illumination area results in a partial field of view in a raw SIM frame, and the stripes in (a) can bring artefacts in SIM reconstruction. Thus, regarding the real sitiuation with optical aberrations, the diffraction intensity from one pinhole with DBBI method is more complete and uniform than the 4:6 dots approximation.

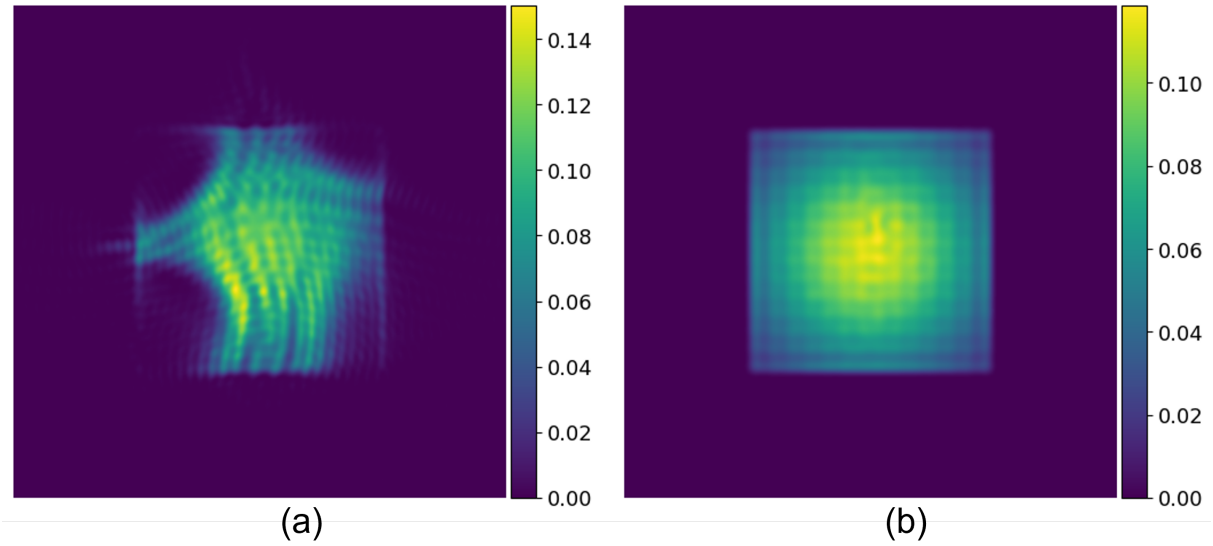


Figure 5-18. The intensity of the light going through one pinhole in the system with the aberration. The light diffracts at the hologram produced using the 4:6 polka dots approximation (a) and DBBI method (b). The hologram in (b) is generated with the constant phase to be zero.

Nervertheless, it is more important to take the interference between diffracted orders into account since this information is used in the reconstruction of super-resolved images from the raw SIM images. In the frequency space, the intensity of the interference between magnitudes (m) passing through pinholes can be written:

$$\begin{aligned} Intensity &= (m_1 + m_2 + m_3)(m_1^* + m_2^* + m_3^*) \\ &= m_1 m_1^* + m_2 m_2^* + m_3 m_3^* + m_1 m_2^* + m_2 m_1^* + m_2 m_3^* + m_3 m_2^* + m_3 m_1^* + m_3 m_1^* \\ &= |m_1|^2 + |m_2|^2 + |m_3|^2 + 2Re[m_1 m_2^*] + 2Re[m_2 m_3^*] + 2Re[m_3 m_1^*] \end{aligned} \quad 5\text{-}27$$

. It can be noted that there is a constant carrier, that is the sum of three squares, and six complex carriers. In the simulation, the diffraction orders of a hologram is calculated as the field of the Fourier transform of the hologram as

$$F_{psf} = FT\left[G e^{i(\pi \times hol + abb)}\right] \quad 5\text{-}28$$

, where G is the Gaussian function, and the result is extened to an area of 4096×4096 pixels to avoid the interference between the edges of the field. The magnitude of the diffraction orders passing through each pinhole and falling on the sample plane can be calculated as

$$m_i = IFT\left[pinhole_i \times F_{psf}\right] \quad 5\text{-}29$$

, where i is the number of the pinhole. The interference between two desired diffraction orders can thus be written from 5-27 as

$$IN_{ab} = 2Re[m_a m_b^*] = 2|m_a||m_b| \cos(\angle m_a - \angle m_b) \quad \text{5-30}$$

, where $\angle$ takes the phase of the individual fields m_a and m_b in the image plane and a and b are numbers of the pinholes (the wanted diffraction orders). These terms are the carriers we need to find in the SIM reconstruction with the cosine term representing the fine structure of the interference terms. In addition, substituting magnitudes calculated by equation 5-29 into equation 5-27, the DC carrier can be calculated as the sum of squares of the individual values as $|m_1|^2 + |m_2|^2 + |m_3|^2$. The three pinholes are shown in Figure 5-19 (a) and are numbered in white. A hologram produced by DBBI including the phase recovery is used to diffract the light into multiple oirders. The interference between diffraction passing through pinhole 1 and 2, pinhole 2 and 3 and pinhole 3 and 1 is shown in (c), (d) and (e), respectively. There are differences between three interferred values, which although are not very obvious. As shown in (b) – (e), there are some dark stripes in the interferred values, which are caused by the periodic modulation in the diffraction intensity (see Figure 5-13) when the phase aberraiton is corrected. The interferred intensities are gird-like since all the interferred light is diffracted from pinholes.

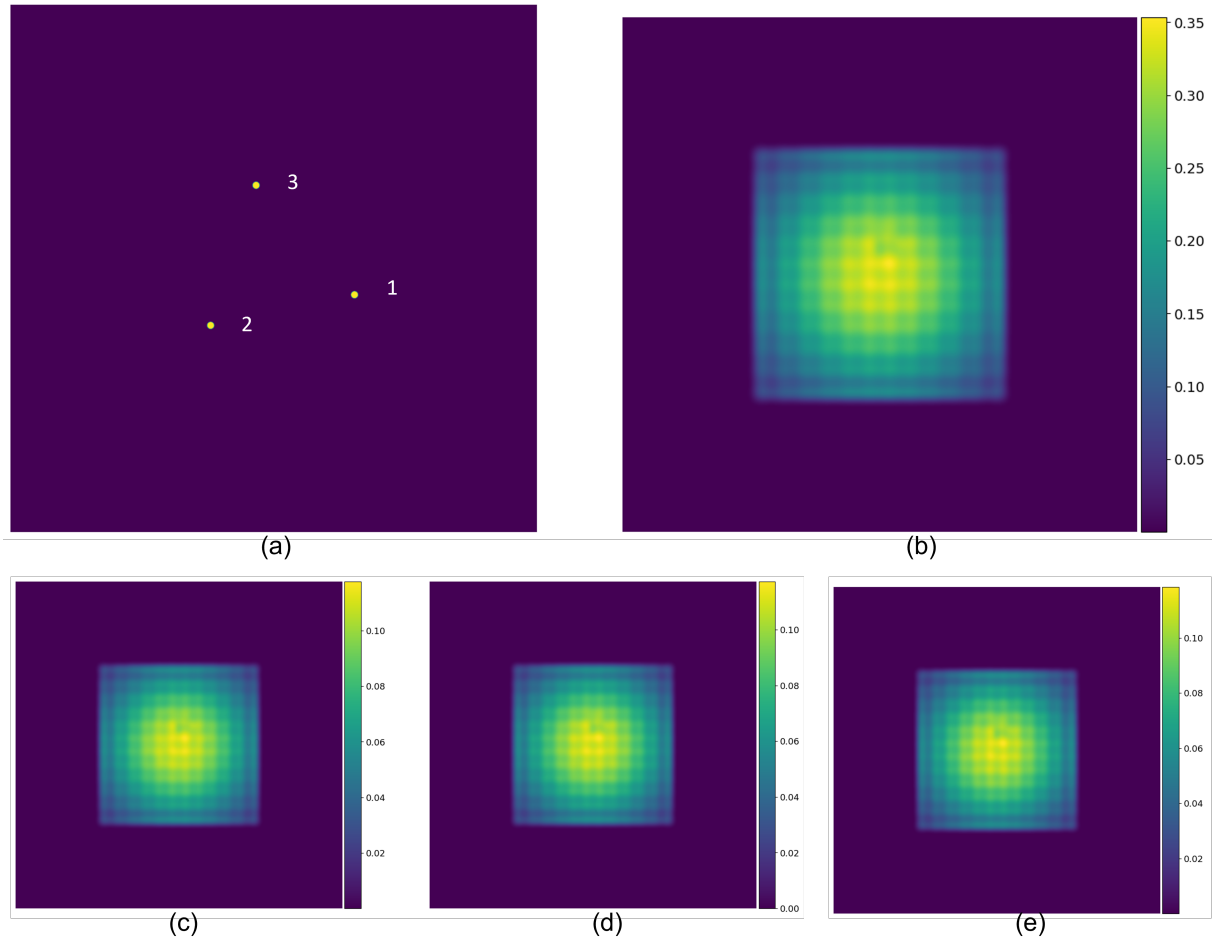


Figure 5-19. The interference between diffraction orders passing through three pinholes when using a hologram generated by the DBBI method. (a) The three pinholes are shown as yellow circles. The number of each pinhole is denoted in white. (b) The DC magnitude generated by the hologram. (c) The magnitude interference between the pinhole 1 and 2 ($|m_1||m_2|$). (d) $|m_2||m_3|$. (e) $|m_3||m_1|$. The unit of colour bars is radians.

While the results of the case with the polka dots approximation, as shown in Figure 5-20, is of distorted shapes, which is caused by the aberration in the SLM's phase.

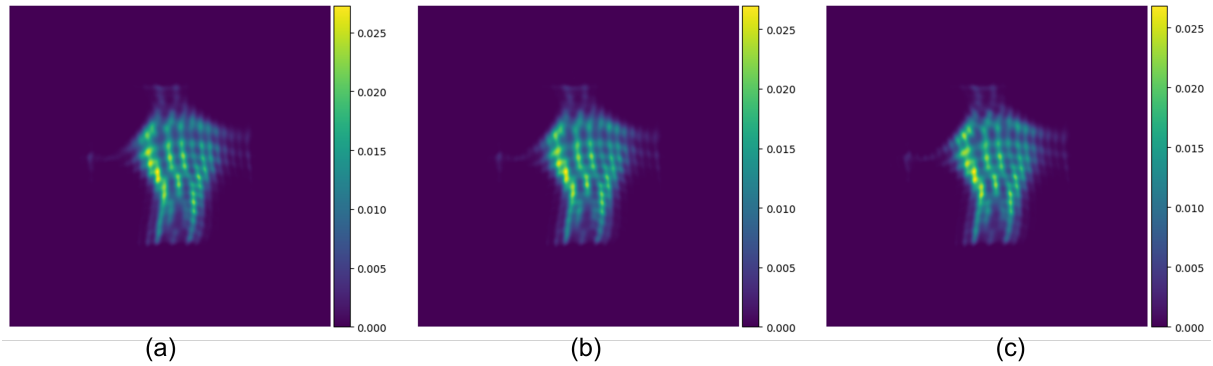


Figure 5-20. (a) The magnitude of the interference between the pinhole 1 and 2 ($|m_1||m_2|$). (b) $|m_2||m_3|$(c) $|m_3||m_1|$. The hologram is produced using 4:6 dots approximation. The unit of colour bars is radians.

In addition, a practical scenario where there is a small misalignment between the three beams and the pinholes is simulated, by introducing some arbitrary shifts in the locations of the three pinholes. A random shift is added in the x (or y) coordinate of a pinhole, with the extent being approximately 83.54% of the radius of a pinhole. For this example, the locations of the three pinholes are set as $(x_1 + 20, y_1 + 12), (x_2 - 7, y_2 + 11), (x_3 - 10, y_3 - 9)$, where (x_i, y_i) is the coordinate of a pinhole, and the size of the mask plane is 4096×4096. As demonstrated in (a), (b) and (c) in Figure 5-21, the three values of the interference between two pinholes changes dramatically compared to results in the top row shown in Figure 5-20, and the covered area is not complete. While the intensities of interfered diffraction orders generated using DBBI approach, as displayed in the second row in Figure 5-21, still preserves the whole range, although some changes caused by the misalignment can be seen. Therefore, DBBI has stronger tolerance in the location of the mask than the dots approximation even though the three beams are misaaligned with the pinholes at 83.54%.

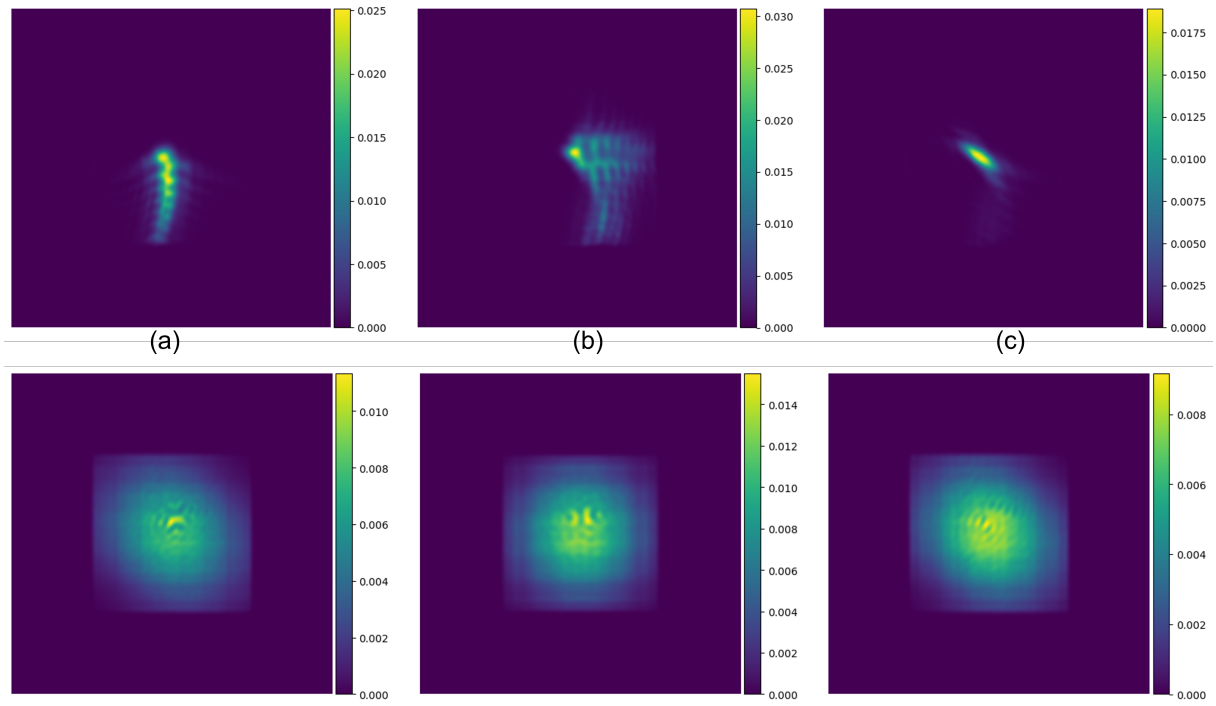


Figure 5-21. The interference between the light transmitting through two pinholes with the misalignment. (a) to (c): The interference between the 1st and the 2nd, the 2nd and the 3rd, and the 3rd and the 1st carriers, respectively, with the hologram generated with the 4:6 dots approximation. (d) to (f): $|m_1||m_2|$, $|m_2||m_3|$ and $|m_3||m_1|$in DBBI case.

5.5. Applying phase stepping in the DBBI method

Although exploiting DBBI to generate holograms can preserve the coverage of the structured illumination and tolerate the imprecise alignment of pinholes, it cannot guarantee a high uniformity. In practice, it is inevitable that other diffraction orders pass through the pinholes, apart form the three orders we want. This leads to non-uniformity of the intensity of the carriers in the sample, and causes incorrect determination of the frequency, amplitude and phase of those carriers when raw SIM frames are reconstructed. However, a strategy of phase stepping can be applied to tackle this issue. In general the n^{th} diffracted order from the hologram will contain $n\times$ the phase of the 1st diffracted order. This means that higher diffracted orders will contain much larger aberration compensations but also if a constant phase ϕ_{con} is added to the field before binarisation then the n^{th} diffracted order will be shifted in phase by $n\phi_0$. The interference between 1st and higher diffracted orders will therefore have a term that varies as $\cos\big((n-1)\phi_{con}\big)$ and can thus be altered by changing the constant phase and further can be averaged out by averaging several intensity patterns while stepping with different ϕ_{con}. This makes the DBBI method further favourable for hologram generation.

In order to make the illumination more even and uniform, we introduced phase stepping in DBBI. Instead of generating just one hologram, multiple holograms with different constant phase are produced. The number of holograms n is also the number of phase steps. For each hologram in the stack, the constant phase is set as $\phi_{con} = \frac{n_{step}\pi}{steps}$, where n_{step} is the number of each hologram, and $steps$ is the entire number of holograms. Note that because of the DC-balancing requirement of the SLM, each hologram will be displayed with both positive and negative contrast, so phase steps in the range π to 2π are also effectively displayed. If the intensity is averaged over all these holograms, then the interference term will only be visible between the 1st and n^{th} diffracted orders when $\frac{(n-1)\pi}{steps} = 2m\pi$, where m is an integer because

$$\sum_{n_{step}=1}^{2*steps} \cos\left((n-1)\frac{n_{step}\pi}{steps}\right) = \begin{cases} \frac{n-1}{steps} \in even & 2*steps \\ otherwise & 0 \end{cases} \qquad 5\text{-}31$$

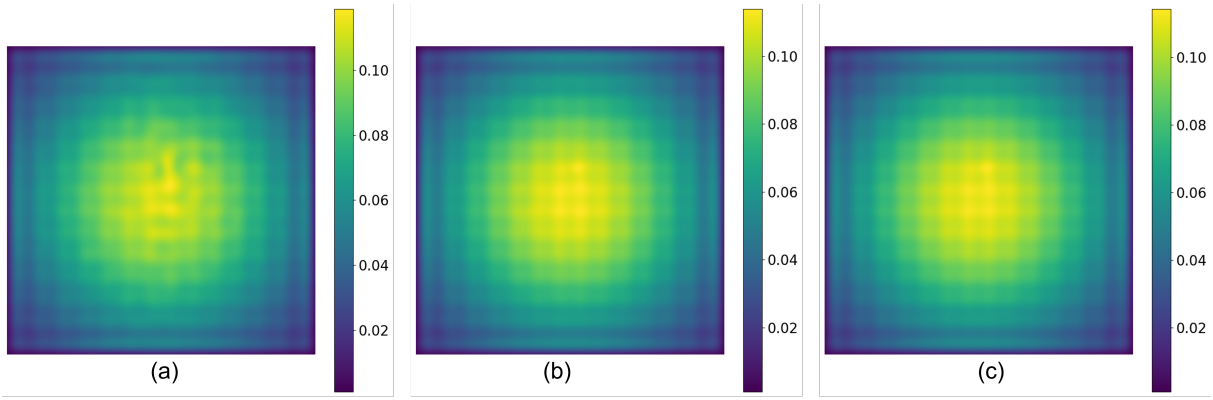


Figure 5-22. The summed illumination through a single pinhole generated using the phase stepping applied in the binarisation of interference between beams with one step (a), 10 steps (b) and 50 steps (c).

The illumination intensities through a single pinhole produced by the diffraction from hologram stacks with different phase steps are summed and displayed in Figure 5-22. Compared to the illumination in the case with only one phase step shown in (a), the results of 10 steps and 50 steps in (b) and (c) are more uniform. Hence, introducing the phase stepping in DBBI can improve the uniformity of the illumination from a pinhole.

However, the difference between improvements using 10 and 50 steps shown in (b) and (c) cannot be judged easily, and the optimal number of phase steps is still mysterious. To analyse the effect of applying phase stepping more accurately, the frequency carriers of the structured illumination are considered. For a given the hologram, the diffraction is calculated as its Fourier transform, and the the diffraction values multiplied with three pinholes are interferred to

illuminate a sample with structured illumination. Seven structured patterns are required to reconstruct one super-resolution image, and the second illumination pattern is shown in Figure 5-23 (a). The illumination intensity is simulated by perfoming inverse Fourier transform of the diffraction passing through three pinholes, which is structured as a hexagonal array of spots, and one small part in it is zoomed-in and shown in (b).

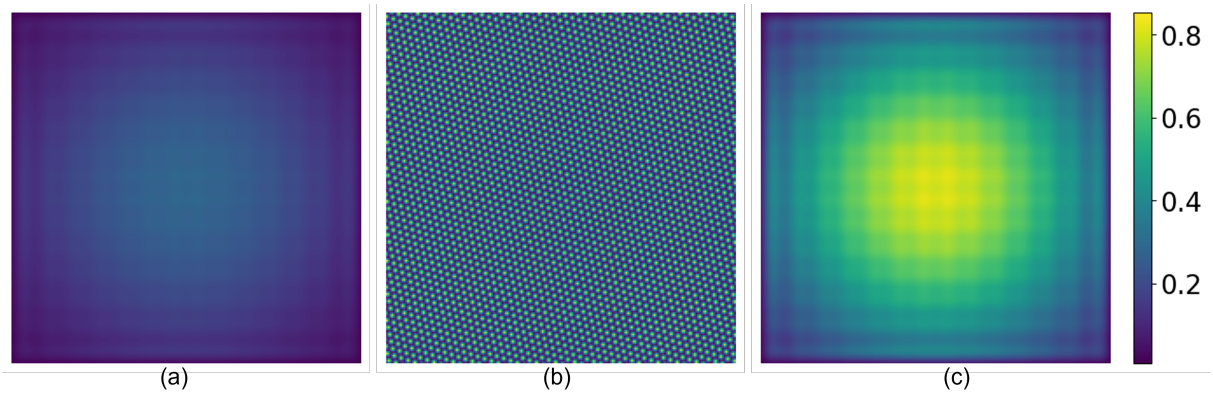


Figure 5-23. The 2nd illumination pattern generated by the interference between diffraction orders passing through the three pinholes. (a) The second structured light pattern. (b) The cropped and zoomed-in part of the 2nd illumination pattern. (c) The 2nd carrier field obtained by the Fourier transform of seven illumination intensities along the axis of seven points.

By applying Fourier transform to the intensities of seven illumination patterns along the seven-point pattern axis, each term corresponds to a carrier component illumination, including the three fundamental frequencies and their conjugates, as well as the zeroth carrier. After the Fourier transform, there are seven fields of carriers. The amplitude of the 2nd carrier field is shown in Figure 5-23 (c).

Two-dimensional Fourier transform can be applied to each carrier field to display the carriers in the frequency domain. The frequencies of the 2nd carrier field are illustrated in Figure 5-24 (a), where seven frequencies can be observed and the second frequency component is dominant. It can also be seen that each carrier field corresponds to a component in the frequency space. The second frequency carrier is cropped and expanded, as shown in (b), displaying both the a clear frquency component and some weak frequencies. All components are produced by the interference between not only the two main diffraction order, but also aliasing higher diffraction orders going though the pinholes.

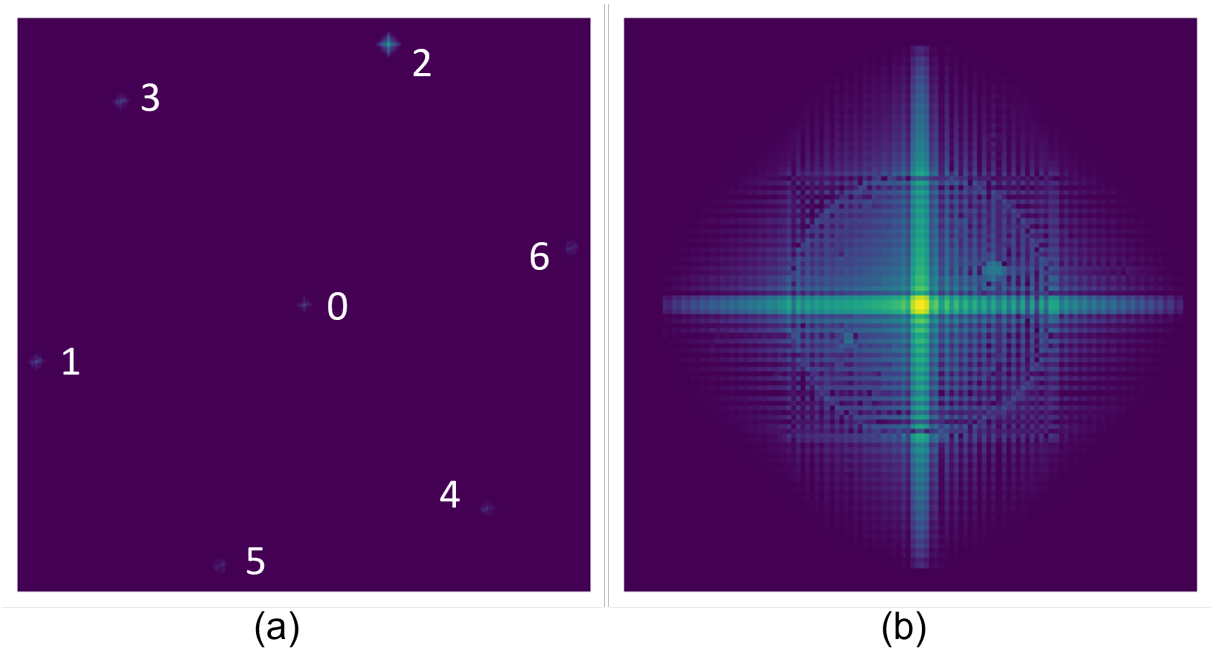


Figure 5-24. (a) The frequency values of the second carrier field. The carrier 2 is dominant here, so it is much brighter than the others. Seven carriers in the frequency domain can be seen and are denotated in white. Log function is used to emphasise the carriers in the image, and the image is cropped. (b) The zoomed-in 2nd frequency carrier.

The amplitude and the phase of each carrier field can be extracted, and the determination of these values is actually significant in processing raw images to get the SIM reconstruction. The amplitude of the illumination determines the magnitude of the field, and is calculated as the absolute value of the illumination intensity in that carrier divided by the intensity of the 0th field. The zeroth carrier field is regarded as the reference, as shown in Figure 5-25 (d). The amplitude of the first, second and the third carriers after dividing by the zeroth value is shown in (a), (b) and (c), respectively, and the amplitude of the third carrier is cropped and zoomed in, as illustrated in (f). Clearly, the amplitudes are not uniform, and some modulation can be seen, although that shown in (a) is flatter than the other two. Additionaly, the phase of the illumination defines the position of the carrier in the frequency plane, and can be obtained by taking the residual phase of the carrier after the tilt due to its nominal carrier position is removed. Similarly, as displayed in (g), (h) and (i), respectively, the residual phase of the 1st, 2nd and 3rd carriers is not uniform. A detailed example is the expanded crop of the phase of the second frequency carrier, as shown in (e). There is no need to calculate the values of both amplitudes and phases of the fourth, fifth and sixth carriers since they are the complex conjugate of the third, second and first carriers, respectively.

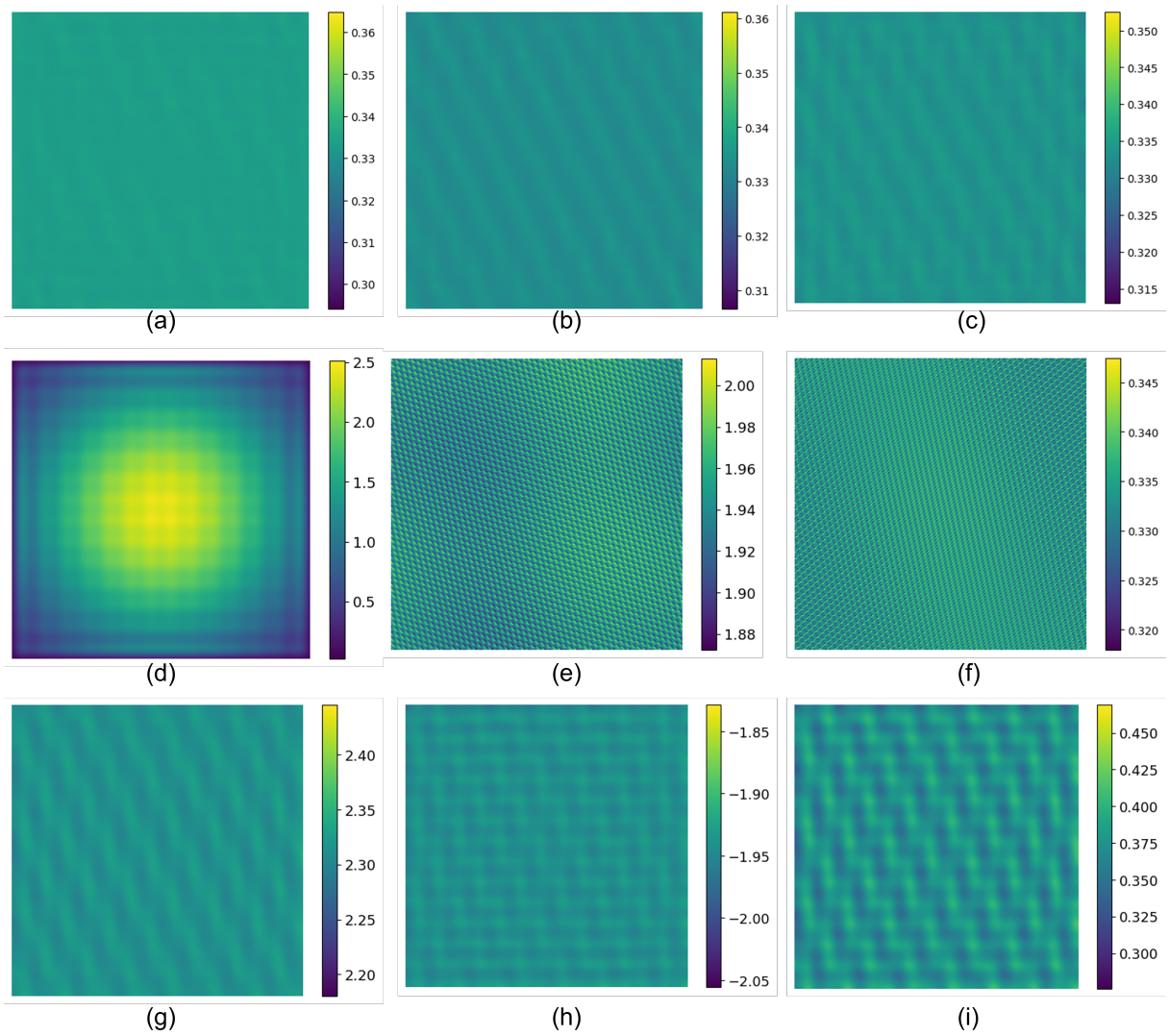


Figure 5-25. Amplitudes and phases of the frequency carriers. (a) to (c): The amplitude of the first, second and third carrier divided by the zeroth frequency, respectively. (d) The intensity of the 0^{th} carrier. (e) The 10x zoomed-in cropped phase of the second carrier (from the centre of (h)). (f) The 10x zoomed-in cropped of the amplitude of the third carrier (see (c)). (g) to (i): The residual phase of the 1^{st}, 2^{nd} and 3^{rd} frequency components, after the expected tilt is removed.

However, these fluctuations in amplitudes and phases can have a bad impact on the SIM reconstruction since they are assumed as constant values across the image field. To address this issue, the intensities in the sample plane are calculated for a stack of holograms with different stepped offset phases to generate the structured illumination. The standard deviation of the amplitude or phase of one of the three frequencies can be calculated from the summation of these illumination images. By introducing multiple holograms with phase steps, the fluctuations are averaged out in the simulation, when the constant phase of each hologram is set as $\frac{n_{step}\pi}{steps}$. In a case without the phase aberration caused by the curved SLM, the DBBI

approach is used to generate holograms with twenty phase steps. As demonstrated in Figure 5-26, the standard deviation σ values of amplitudes relative to the zero order and residual phases for the three frequency components reduces as the number of phase steps increases, although there are some peaks. Three peaks appear in both the amplitude and phase of the first carrier with a period of three steps. Both σ of amplitudes and phases of three carriers decline with the phase steps and thus result in a similar and low value compared to that at the first step, especially for the amplitudes. Apparently, the phase stepping has an obvious positive effect on determining the two parameters of frequencies. In fact, as shown in Figure 5-26, taking ten phase steps is enough to reduce the fluctuation in the relative amplitudes and the residual since the values at the tenth step cannot be reduced dramatically with the additional ten phase steps. Considering the efforts in designing more holograms and the time for transferring them to the SLM, ten phase steps is a decent choice.

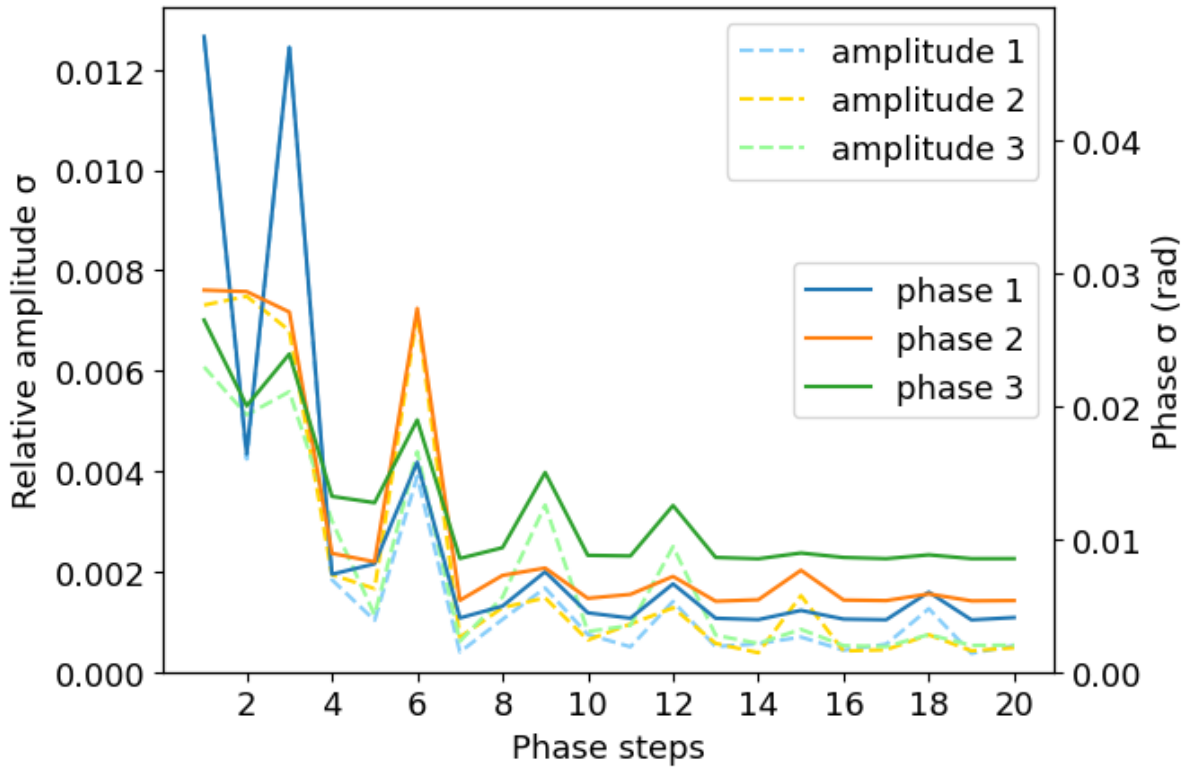


Figure 5-26. The standard deviation σ of the amplitudes and phases of three frequency components with phase steps in an anerration-free situation. The range of the phase steps is from one to twenty. The holograms are produced using the DBBI method.

To discuss the extent of the decrease in the standard deviation of parameters of the frequency components contributed to using the phase stepping, a fractional decrease D_{20} is calculated as

$$D_{20} = \frac{\sigma_{20} - \sigma_1}{\sigma_1} \quad 5\text{-}32$$

, where σ_1 is the initial standard deviation and σ_{20} is the value when twenty phase steps are taken. Initial values and the decrease ratios of all parameters are shown in Table 5-3. The decrease in these parameters is significant, especially in the amplitudes, with D_{20} being larger than 90% for all carriers. Therefore, by using multiple holograms with the phase stepping, the uniformity of phases and amplitudes can be refined, leading to a higher accuracy in determining parameters of the illumination in the frequency domain, and the SIM reconstruction can thus be improved by reducing artefacts. The values of three carriers are different since they are generated by the interference between the main diffraction orders and different higher orders. We believe that if the orientation of holograms (and the pinholes) is changed, the values of three components will be different.

standard deviation σ without aberration or correction	Carrier 1		Carrier 2		Carrier 3	
	1 step	D_{20}	1 step	D_{20}	1 step	D_{20}
Amplitude relative to 0th order	0.0126	-95.9%	0.0073	-93.5%	0.0061	-91.2%
Residual phase (radians)	0.0477	-95.1%	0.0287	-81.3%	0.0264	-67.8%

Table 5-3. The standard deviation in amplitudes and phases of three carriers and the fractional decrease ratio with 20 phase steps (D_{20}). DBBI is utilised to design holograms and there is no phase aberration in the system.

Furthermore, the realistic case with the aberration caused by the SLM in our set-up is simulated. The DBBI approach is used to design hologram with the phase steps to be twenty and the phase correction, and the aberration practically measured from the SLM is added in displaying holograms. Compared to ideal aberration-free situations, the realistic case demonstrated in Figure 5-27 also exhibits the decrease in values of the standard deviation in relative amplitudes and residual phases of the three frequency carriers, but the peaks in these lines are not noticeable after the sixth phase step.

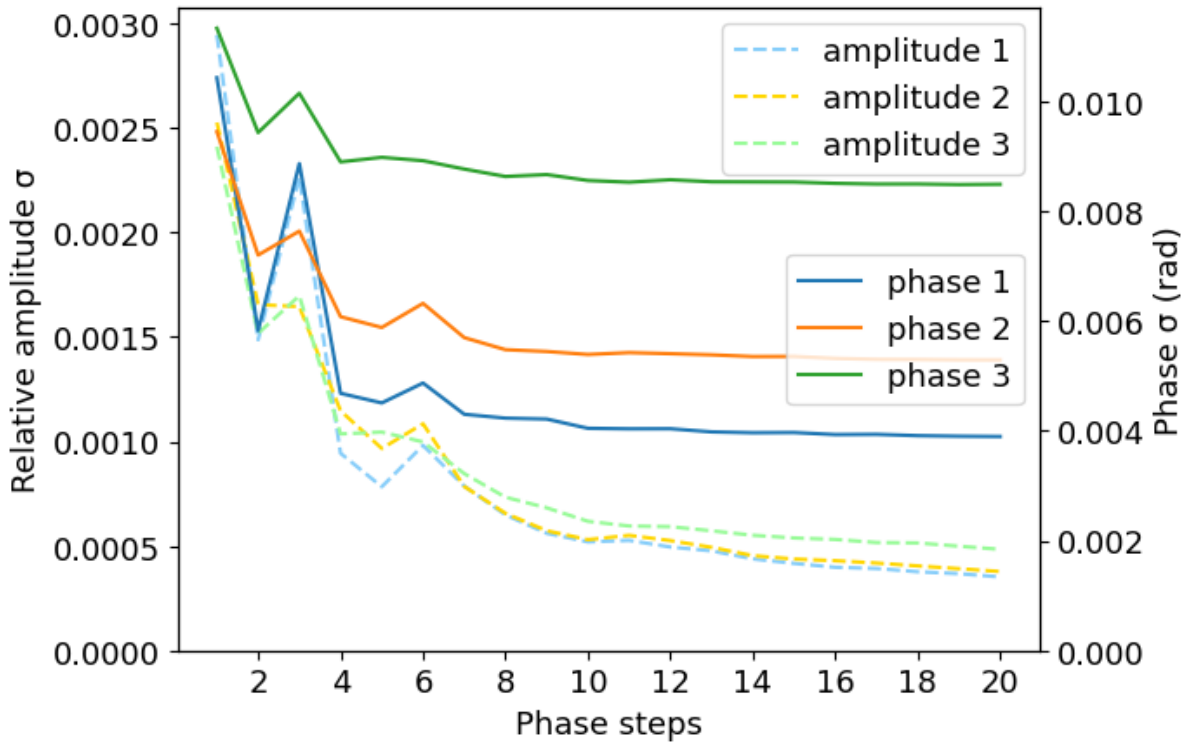


Figure 5-27. The standard deviation σ of the relative amplitudes and residual phases of three carriers with 20 phase steps with the aberration obtained from the optical setup. The method to generate holograms is taken as the DBBI with the phase correction.

In fact, as shown in Table 5-4, the σ values of both the amplitudes and the phases at the first step are lower than that in aberration-free case (see Table 5-3), means that the DBBI method with the phase correction can provide more smooth parameters when the aberration is recovered than the case without aberration or the correction. Additionally, all D_{20} values shown in Table 5-4 demonstrating the DBBI's capability to further reduce the fluctuations in the parameters.

standard deviation σ with aberration and correction 100μm pinhole	Carrier 1		Carrier 2		Carrier 3	
	1 step	D_{20}	1 step	D_{20}	1 step	D_{20}
Amplitude relative to 0th order	0.0029	-88.1%	0.0025	-85.1%	0.0024	-80.0%
Residual phase (radians)	0.0104	-62.7%	0.0094	-44.0%	0.0113	-25.1%

Table 5-4. The standard deviation σ and D_{20} in amplitudes and phases of three carriers with holograms generated by DBBI method with the phase correction.

Therefore, the phase stepping can be used to increase the consistency of the three carriers, whether there is an optical aberration in the system or not. In practice, we implemented this approach by displaying the hologram stack during the exposure time of a camera for capturing one raw image, the sample can thus be illuminated by the interference of uniform light. For example, the frequency of displaying hologram of the SLM (Forth dimension displays, R11, 2K) in our system can be as high as 10000 frame per second, which make the phase stepping feasible with a large number of phase steps (although ten-phase-stepping is good enough for suppressing fluctuations in the parameters). Given a camera exposure time for one raw SIM frame as low as 10s of milliseconds, introducing the phase stepping would not affect the speed of capturing images.

5.6. Pinhole size

Since the low diffraction efficiency and the uneven illumination are both mainly caused by the large extension of the diffracted orders, utilising larger pinholes would seem to be an alternative solution for coping with the aberration. However, large pinholes give rise of more unwanted diffraction orders passing through the pinhole, and the resultant non-uniform illumination cannot be solved in the polka dots approximation or the optimisation of binary search. In contrast, by employing the DBBI method, the pinhole can be even smaller since the corrected diffraction beam is diffraction limited, and the diffraction intensity are more contributed by the three wanted beams.

The amplitude of the diffraction orders emitting at the mask plane with using three methods to generate holograms is simulated, respectively, and is shown in Figure 5-28. When adapting the 4:6 dots approximation, compared to our realistic case where the pinhole radius is $100\ \mu m$ (see a), the aberrated diffraction orders can go through pinholes completely when the radius is enlarged twice (see b), while the diffraction orders shown in (c) are significantly blocked by small pinholes with the radius as $50\ \mu m$. However, if the DBBI approach with the phase correction is used, even though the phase step is set as one and the constant phase is zero, as demonstrated in (d) – (f), the diffraction orders can always go through the pinhole entirely, regardless the pinhole size. It can be seen that higher orders in the DBBI case are much weaker than results produced with the 4:6 dots approximation (see the first row) because the applied phase recovery makes light mainly focused into the three desired orders.

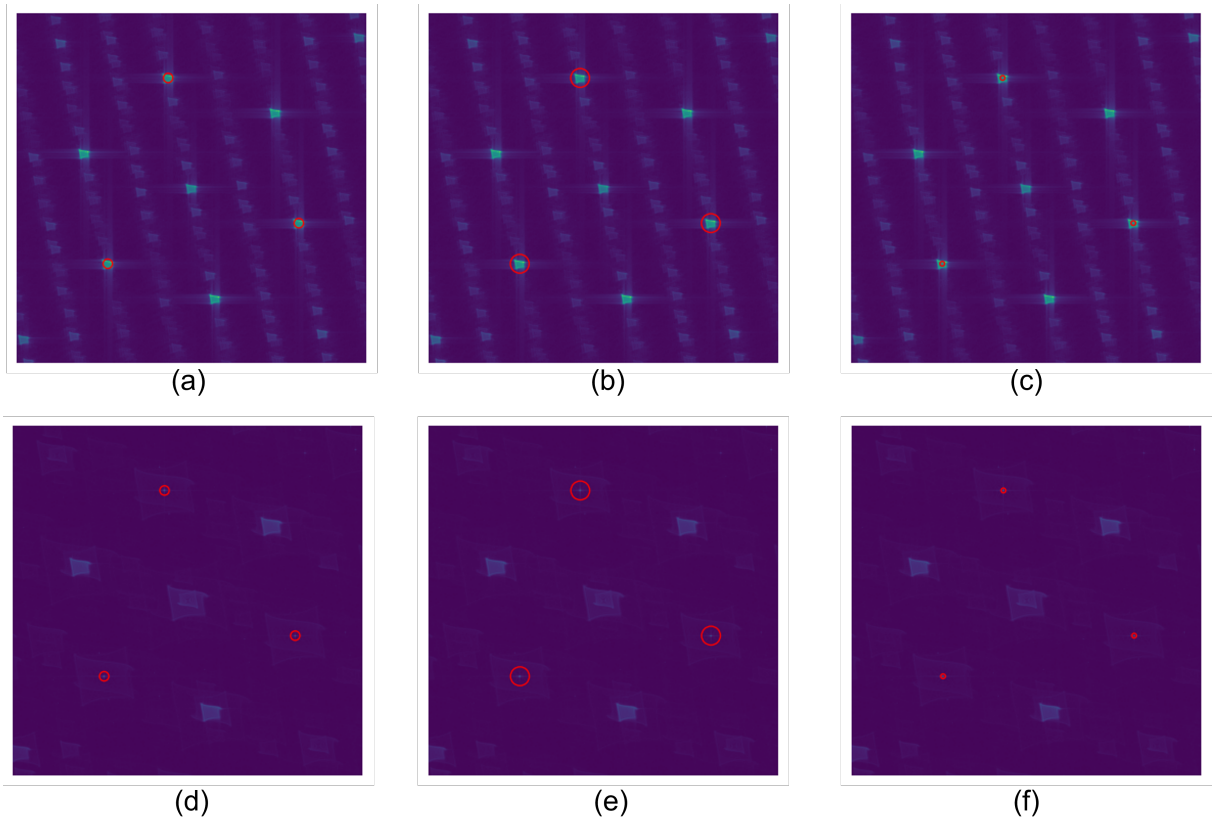


Figure 5-28. The amplitude of the light orders diffracted by holograms generated by the 4:6 dots approximation (top row) and the DBBI method (bottom row) with different sizes of pinholes. The amplitude of the diffraction orders with utilising a pinhole with the radius as (a) 100 μm (which is the same parameter in our practical mask), (b) 200 μm and (c) $50\mu m$. (d) – (f) are diffracted orders of a hologram generated by the DBBI approach with the constant phase to be zero, passing through pinholes with different radii.

The values of the interference between two fields diffracted from a pinhole are shown in Figure 5-29, with the radius ($200\,\mu m$) of pinholes being twice than our real situation. Since the performance of holograms designed by the 4:6 polka dots and the optimisation by binary search are similar, we only focus on the comparison between the polka-dots approximation and DBBI in this figure. As demonstrated in the top row, when using a mask with the twice pinholes size and the 4:6 polka dots, the field of view of the interference are entire since extended diffraction orders can pass through pinholes completely, although there are some wavy stripes in the interfered light caused by aliasing higher orders. When DBBI is applied with the large pinholes size, as shown in the bottom row, the coverage of the interference is complete. Note the dark fringes appearing in in figures (d) to (f) that are the binarisation artefact shown in Figure 5-13.However, only according to the images of the interference, it is difficult to determine which method of designing holograms is superior when large pinholes are applied.

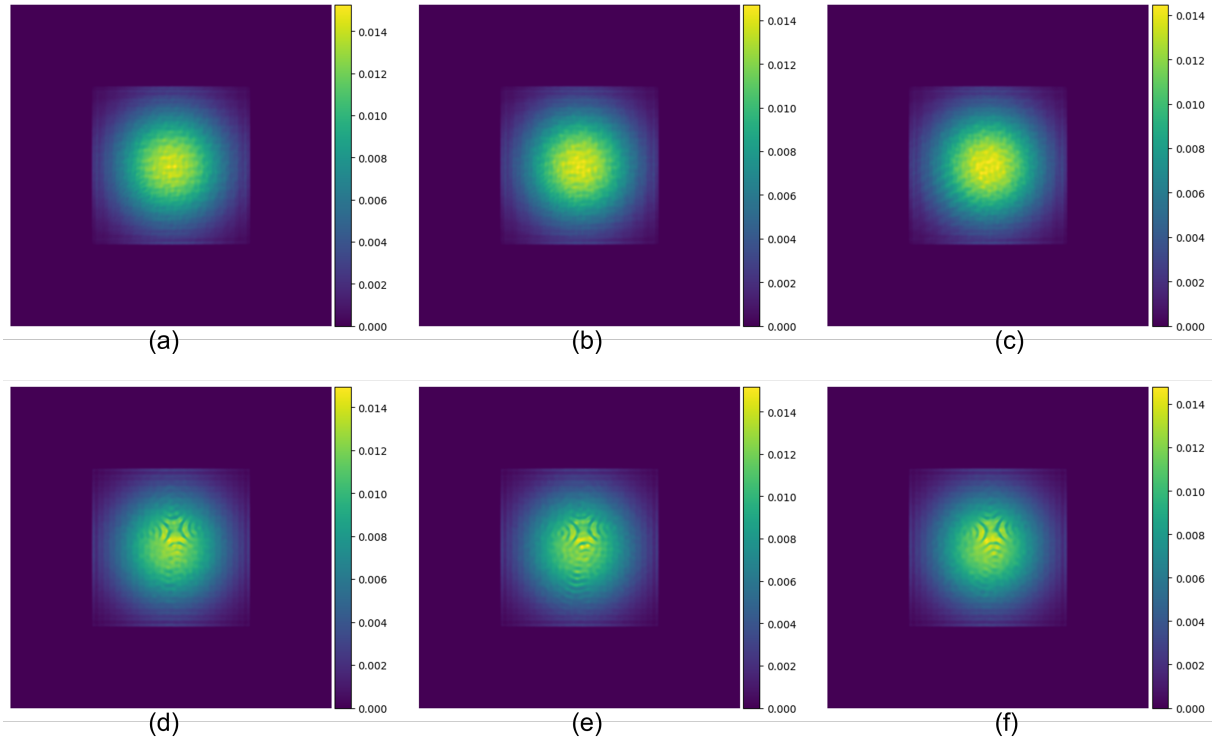


Figure 5-29. The interference between diffraction passing through (a) the 1st and the 2nd, (b) the 2nd and the 3rd, and (c) the 3rd and 1st with a pinhole radius of 200 μm and the hologram generation as the 4:6 polka dots approximation. (d) to (f): IN_{12}, IN_{23} and IN_{31}, respectively, with the 200 μm radius but exploiting DBBI.

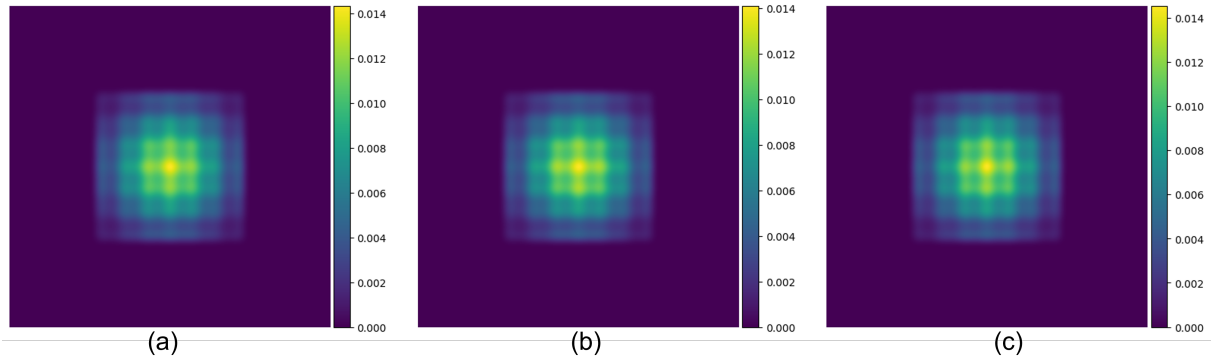


Figure 5-30. The interference between diffraction passing through two pinholes with the radius as as 50 μm and the hologram design as DBBI. (a) IN_{12}. (b) IN_{23}. (c) IN_{31}. The holograms are generated by DBBI.

In contrast, if DBBI is used, the pinhole size can be even smaller since the beams are corrected to diffraction limited, and many undesired frequencies can be excluded in the carriers. As demonstrated in Figure 5-30, as the radius (50 μm) of the pinhole is set as the half size, a grid-like modulation are shown in the interference between light diffracted from two pinholes, instead of the artefacts displayed in Figure 5-19 (c) – (e) and in Figure 5-29, with the pinhole

radius as $100\ \mu m$ and $200\ \mu m$, respectively. The interference value is mainly modulated by the interference from the edges of the pinholes. However, exploiting a very small pinhole radius cannot offer a high tolerance in the different locations of the three pinholes and diffraction orders.

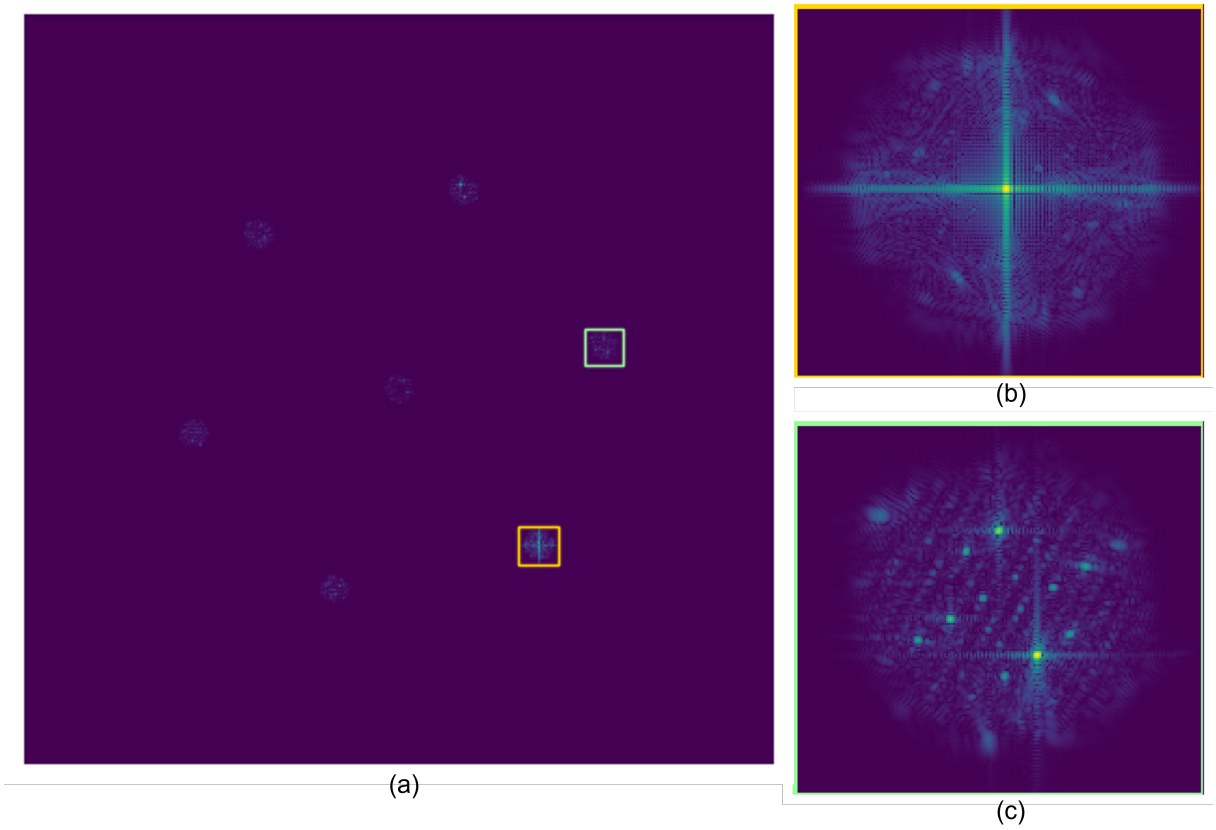


Figure 5-31. Frequency carriers with 200 μm pinholes. Holograms are produced by the 4:6 polka dots approximation. (a) The frequency values of the 3rd carrier field. Seven carriers in the frequency domain can be seen and the 3rd (denoted in the orange frame) is dominant. Log function is used to emphasise the carriers in the image. (b) The zoomed-in 3rd frequency carrier. (c) The zoomed-in 1st frequency carrier, which is framed in a green rectangular.

In fact, considering about the carriers in frequency domain, the polka-dots method with a large pinhole radius is still not favourable to obtain accurate parameters. As displayed in Figure 5-31 (a), the seven carriers in the polka-dots case with $200\ \mu m$ pinholes have larger ranges because of the larger pinhole size than Figure 5-24 (a), including more unwanted frequency orders. The main frequency is the third carrier, which is frame in an orange rectangular, while one of others, the first carrier is denoted by a green frame. As shown in Figure 5-31 (b), apart from the desired frequency, there are many components in the third carrier. Additionally, there are many frequencies in other carriers, as shown in the first carrier in (c). All undesired frequency

components presented in both the main carriers and others will interfere, which can impede the determination of the amplitude and the phase of this carrier and then has a negative effect in SIM reconstruction.

Since 100 μm pinholes do not let all the light through when a typical aberration is present, the DBBI with aberration correction is exploited to get all the light through a 100 μm (or even smaller) pinhole, while the 4:6 polka dots approach is cooperated with a large radius (200 μm) of a pinhole. In terms of the standard deviation σ in the relative amplitudes and the residual phases of three frequency components, the 4:6 dots approximation and the DBBI method with the aberration correction are compared when the system exist the aberration caused by the SLM.

σ at 1 step / Pinhole radius (μm)	A_1	A_2	A_3	φ_1(rad)	φ_2(rad)	φ_3(rad)
200 μm Polka dot	0.0070	0.0054	0.0069	0.0221	0.0164	0.0204
200 μm DBBI	0.0062	0.0054	0.0049	0.0211	0.0170	0.0165
100 μm DBBI	0.0029	0.0025	0.0024	0.0104	0.0094	0.0113
50 μm DBBI	0.0019	0.0015	0.0018	0.0054	0.0056	0.0066

Table 5-5. The standard deviation σ in amplitudes and phases of three carriers at 1 step using different pinhole sizes. A and φ is the symbol of the amplitude and the phase of a frequency carrier, respectively. DBBI is utilised to design holograms and correct the phase aberration.

In all these cases, the three desired diffracted orders can pass through the pinholes, and σ values are shown in Table 5-5. When the pinholes radius is large (200 μm), the σ values with the phase correction (see the second row) are approximately the same as that with a uncorrected hologram (see row 1). Reducing the pinhole size reduces these phase and amplitude variations approximately by the linear dimension of the pinhole.

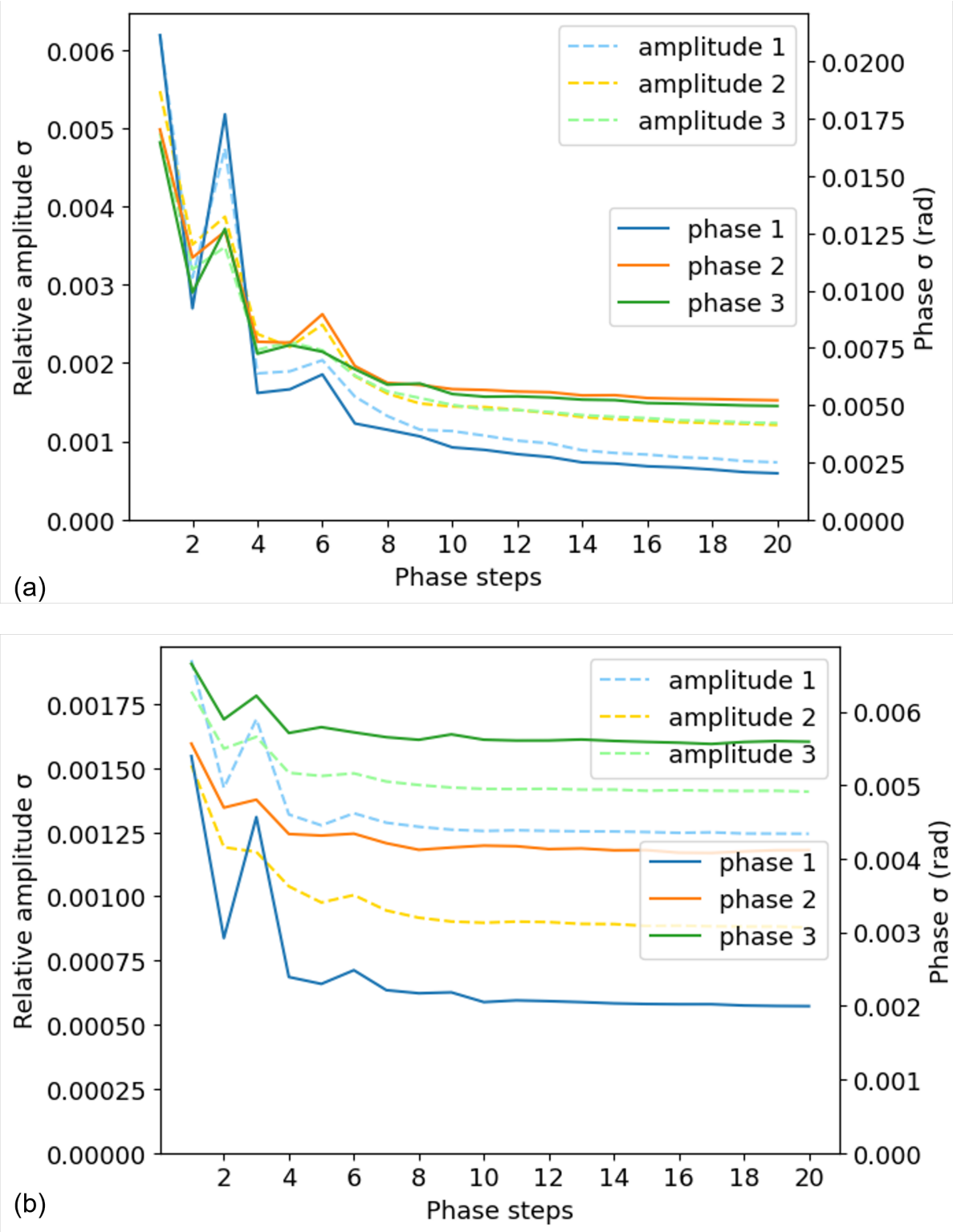


Figure 5-32. Variations in the relative phases and residual phases with 20 phase steps when the radius is (a) 200 μm and (b) 50 μm. The method to generate holograms is taken as the DBBI with the phase correction.

The phase stepping with twenty steps is added into the DBBI method for cases with the 200 μm and the 50 μm pinhole radius, respectively. The variations in the amplitudes and phases

are presented in Figure 5-32. Similar to Figure 5-27, the variations in the parameters of three carriers decline with phase steps although there are relatively large peaks at 3 steps and 6 steps. Simulations are shown for up to 20 phase steps, though the performance at only 10 phase steps is roughly similar.

The specific σ values are shown in Table 5-6. Phase stepping the holograms reduces the phase variations for each pinhole size to about the same level. The effectiveness of the phase stepping is therefore more apparent when using larger pinholes. Phase stepping the holograms also reduces the amplitude variations for each pinhole size, but the lowest levels achieved appear to be for the 100 μm pinhole.

σ at 20 steps / Pinhole radius (μm)	A_1	A_2	A_3	φ_1(rad)	φ_2(rad)	φ_3(rad)
200	0.0007	0.0012	0.0012	0.0020	0.0052	0.0050
100	0.0003	0.0004	0.0005	0.0039	0.0053	0.0085
50	0.0012	0.0009	0.0014	0.0020	0.0041	0.0055

Table 5-6. The standard deviation σ in amplitudes and phases of three carriers at 20 steps using different pinholes sizes. A and φ is the symbol of the amplitude and the phase of a frequency carrier, respectively. DBBI is utilised to design holograms and recover the aberrated phase.

Therefore, The DBBI with the aberration correction and phase stepping permit the use of smaller pinholes and reduces the amplitude and phase variations that arise from interference with aliased diffraction orders resulting from the pixelation of the SLM.

5.7. Chapter summary

In this chapter, a mini review is given at the beginning to introduce SIM systems using an SLM to generate structured illumination. After explaining the frequencies enclosed in a hexagonal hologram, three methods for generating holograms exploited in a HexSIM system are introduced. The hologram can be directly generated by approximating the polka dots and setting the black and white with the same area. The second way (DBBI) is to binarising the interference between the three wanted beams where the constant phase can be adjusted, and a correction phase can be added for recovering the phase aberration in the system. The third

method is to optimise a random hologram according to the intensity of three beams as the target.

The three methods are compared in multiple aspects to determine a superior way to generate holograms. In a system with zero aberration, all three methods can achieve the high diffraction intensity, although the ratio of the positive and negative area is required to be set as 4:6 and there are fluctuations with a period of $\frac{\pi}{3}$ in the DBBI result as the constant phase changes. When the more realistic scenarios which include the aberration caused by the SLM are considered, the diffraction intensity of three methods in a system with the phase aberration can be the same. With the aberration, the fluctuation in the diffraction intensity with the constant phase is small, and the maximum can be higher than results of the other two methods. Furthermore, utilising DBBI can provide a uniform and complete diffraction intensity since the phase aberration is corrected, while other two methods show incomplete coverage of the field of view.

The speed of the polka dots approximation is the highest among three methods, and a hologram of 500×500 pixels can be generated in 0.028s. The DBBI is also fast (0.047s), while the optimisation by the binary search is slow, which takes about 38.6s to produce one hologram. The performance of the 4:6 polka dots and the optimisation without extension for the phase correction and phase stepping are approximately the same. The optimisation by binary search is not refined and implemented in realistic cases because the time consumption of designing holograms is too high.

The illumination at the sample is discussed in cases using the 4:6 polka dots and the DBBI with the phase correction, in terms of both the diffraction intensity passing through one pinhole and the interference between the light travelling through two pinholes. The illumination of the dots approximation shows an incomplete field, while that of the DBBI is complete since the (two) desired diffracted order (orders) can pass through the (two) pinholes. Additionally, the DBBI with the phase recovery has a high tolerance of the alignment between the three diffracted orders and pinholes. When there is a misalignment that is approximately 83.54% of the pinhole radius (100 μm), the interfered values of the case with the 4:6 dots hologram almost lose the field of view, while the DBBI case keeps the complete interfered illumination.

Residual artefacts in SIM reconstruction can be attributed to phase and amplitude variations in the generated frequency carriers. To arrange out the variations and thus to determine

accurate values for a SIM reconstruction, the phase stepping is introduced into the DBBI method with the phase correction. In both the situations with and without the phase aberration, the standard deviation values of the amplitudes and phases decrease with phase steps.

In fact, the low diffraction efficiency and the incomplete illumination are resulted from that diffracted light cannot pass through 100 μm when a typical aberration is present. Apart from applying the DBBI with the aberration correction, using the 4:6 dots approximation with a larger pinhole radius (200 μm) can make the light through pinholes entirely. However, aliased higher diffracted orders passing through the large pinholes can cause relatively high carrier phase and amplitude variations. By contrast, the pinhole size can be minimised to (50 μm) if the DBBI with the aberration correction is exploited. Decreasing the pinhole size approximately reduces these phase and amplitude variations in proportion to the pinhole's linear dimension. With applying the DBBI with the phase recovery and twenty phase steppes, the resultant phase variations for each pinhole size are approximately the same. Hence, the effectiveness of the phase stepping is more significant for cases with larger pinholes. The phase stepping also reduces the variations of amplitudes, however, the minimum values appear to be in the 100 μm situations. The variations at ten steps are roughly the same as that at twenty steps, means that the number of phase steps can be chosen as ten in practice. The DBBI aberration correction and phase stepping allow for the use of smaller pinholes and minimise the amplitude and phase variations caused by interference with aliased diffracted orders due to the pixelation of the SLM.

To sum up, the polka approximation is the fastest and the most direct method but cannot provide any correction or improve the entirety and uniformity of the structured illumination. Theoretically, the method of optimisation by binary search can be adapted to achieve the high diffraction intensity and even correct the aberration in the system. However, the low speed of optimising the hologram pixel by pixel makes this method impractical. Utilising a mask with larger pinholes to let extended diffraction orders go through can address the entirety of the illumination, nevertheless, can increase variations in relative amplitudes and residual phases of frequency carriers, thus impede the determination of frequency carriers of the illumination in SIM reconstruction. The DBBI method with aberration correction and ten (or twenty) phase steps is the most favourable and practical way to generate holograms, especially for a system with some optical aberrations, which is actually quite common. The hologram generated by DBBI with phase steps can produce the highest diffraction intensity and also the most uniform and complete illumination projected on the sample plane, with the pinholes being not required to be enlarged.

Chapter 6. Implementation and measurements

Implementing SIM imaging has always been considered a complex task due to the intricate alignment requirements and need for comprehensive control of the equipment. This chapter focuses on imaging samples with hexagonal structured illumination (HexSIM) and presents this particular implementation in terms of hardware and software. The purpose of this chapter is to introduce the system and platforms we use and to show the superior results obtained with applying the DBBI method with phase correction and phase stepping. At the beginning of the chapter, the practical system is introduced, which features easy and remote alignment capabilities through the use of computer controlled linear actuators and cameras to remotely monitor alignment. All hardware is controlled from Python based on the ScopeFoundry platform. This program integrates functions controlling the SLM, a DAQ (data acquisition) box, lasers, a z stage and two cameras; one is used for in-situ phase correction and the other for acquiring raw SIM image frames. The implementation of in-situ phase recovery is discussed before detailing the synchronisation between displaying gratings (holograms) on the SLM and the camera's exposure. The specific repertoire (the arrangement of orders and duration of displaying holograms) loaded on the SLM is explained. The results of the in-situ phase correction are presented, showing diffraction limited diffracted orders that can pass entirely through the pinholes on the spatial mask. Given the phase recovery, the wide-field images of fluorescent samples present a more complete field of view (the visible area in a captured image) than that without the correction. The reconstructed image in the case using the DBBI method is compared with the case with phase correction and the situation with both phase recovery and phase stepping in terms of the spectrum. Obviously, the scenario without phase correction or phase stepping provides the vaguest spectrum. The uniformity of the brightness and the phase at each hologram in reconstructed images are compared between the case with only phase correction and that with both phase correction and stepping. Apparently, the DBBI method with phase recovery and stepping is the most favourable approach to obtain super-resolved images with fewest hot pixels and weakest fluctuations in the phase at each hologram.

6.1. Optical system and platforms

Building on the HexSIM system first outlined in Chapter 2 (first introduced by Gong, Guo and Neil [38]), we developed a system to conduct SIM imaging using hexagonal illumination patterns with two-colour laser light. This particular SLM HexSIM system was designed to minimise difficulty when aligning the optical path and to reduce the inconvenience involved in

controlling the various pieces of equipment, as will be explored in this chapter. In addition, a bespoke Python platform: HexSIM ScopeFoundry is developed to control the hardware and visualise the SIM images.

6.1.1. Optical layout of the HexSIM system and remote control

Figure 6-1 shows a system diagram for the SLM HexSIM system, including the excitation and emission light paths.

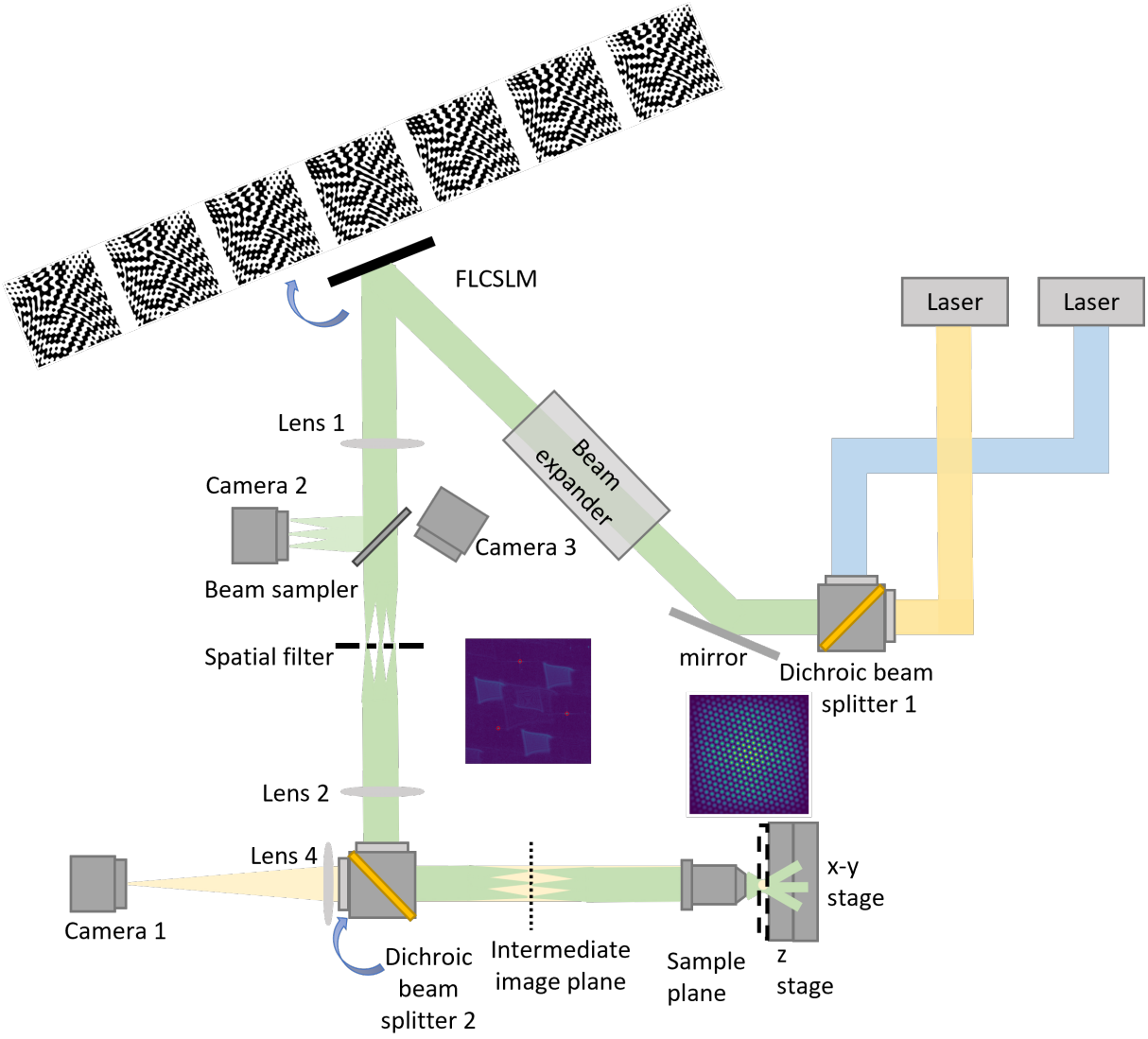


Figure 6-1. Schematic of the HexSIM system. Only main equipment is shown in the image, some parts such as relay lenses, the microscope frame and some supporting equipment (e.g. a data acquisition box) are omitted. Seven holograms with phase correction are presented.

As can be seen in Figure 6-1, two separate laser illumination beams (Oxxius, LCX-561L and LBX-488) are co-aligned and combined in dichroic beam splitter 1. The laser light beams are

first directed through a linear polariser, then expanded using a beam expander that includes a pinhole spatial filter, and finally directed onto the screen of the ferroelectric liquid crystal spatial light modulator (FLC SLM). (Forth dimension display, R11, 2K). Since the holograms displaying the hexagonal structures as described in chapter 5 are loaded onto and subsequently displayed on the SLM, the incident laser light will diffract into multiple orders. Lens 1 performs a Fourier transform to produce the diffraction orders, focusing them onto the spatial filter, namely, the pinhole mask. In addition, a beam sampler (Thorlabs, BSF10-A) is placed between the mask and lens 1 in order to direct ~10% of the diffracted light to camera 2 (Thorlabs, CS2100M-USB), which is used for the phase recovery implementation (previously discussed in chapter 4). For the main imaging path, only the three desired beams can pass through the mask, with the zeroth order, the three conjugates and other higher orders being blocked. As shown in the figure, the three diffracted orders are corrected to be diffraction limited. The spatial mask is installed in a filter wheel, along with other masks with various different pinhole seperations. It is thus possible to characterise the performance of different masks easily and conveniently. The three diffracted beams are then collimated and go through dichroic splitter 2, before an image of the SLM plane is focused on the intermediate image plane by the combination of lens 1 and lens 2. Then, the beams are focused by microscope (Olympus IX51) internal tube lens into the pupil plane of the objective lens (Olympus UPlanFl 60x/1.25 Oil), which then projects the structured illumination generated by the interference between them onto the sample plane. The illumination excites fluorescence emission the sample, this emitted light passes back though the dichroic and finally reaches the camera (Hamamatsu, ORCA-Fusion C14440), relayed by a custom x1 magnification relay lens [99], where raw SIM images can be captured. The position of the sample stage is controlled by a piezo z-stage (Mad City Labs INC., piezo amplifier Nano-Drive® C) in the axial direction and a motorised scanning stage (Märzhäuser Wetzlar GmbH & Co.KG, SCAN IM) in the lateral directions.

There are a number of features we have implemented in our SLM-SIM system which help to make alignment easier and more accurate. In an SLM-SIM system, the SLM diffracted orders must be aligned with the mask plane, and the mask also needs to be aligned with the back focal plane of the objective. To tackle these challenging alignments, the SLM is mounted on a tip-tilt stage, with the tip-tilt angles controlled by two piezo inertia motor actuators (Thorlabs, PIAK10). Similarly, the mounting tip-tilt angles of the dichroic beam splitter 2 located between camera 1 and the microscope objective are also controlled using two similar actuators of the same model. Each actuator has a travel range of 10 mm with a step size of 10-30 nm. All four actuators are controlled by a Four-channel K-Cube piezo Inertia motor controller (Thorlabs,

KIM 101). Both SLM and dichroic alignment can be carried out remotely. The piezo stage controlling the SLM, directs the diffracted the light into the three pinholes on the mask, as observed by camera 3. While the adjustable dichroic splitter allows the alignment of the diffracted orders from the pinhole mask with the objective pupil. In this way there both paths are aligned to the pinhole mask separately, which then requires no lateral alignment. The alignment can thus be carried out simply and remotely.

Both lasers used in the HexSIM system are class 3B, which could pose a hazard to the eyes and skin of a researcher. Therefore, for the sake of both safety and convenience, the entire system is enclosed in a box, excluding the microscope frame. It follows then that all equipment has to be aligned and controlled remotely. To observe the alignment of the hologram and the mask plane, camera 3 (XIMEA, xiQ) was installed to observe the mask plane, which was illuminated using a LED lamp that is omitted in the figure (Thorlabs, MWWHL3). As shown in Figure 6-2, given the image of the mask plane with the illumination from the LED, the position and the angle of the diffracted orders can be observed. The figure demonstrates a situation where the alignment between beams and the mask requires some adjustment. The location of the diffraction orders can be translated using the controller placed outside the safety box controlling the two actuators to adjust the SLM angle. The angular rotation of the diffracted spots to the pinhole mask, however, is changed by adding a rotation when designing the hologram.

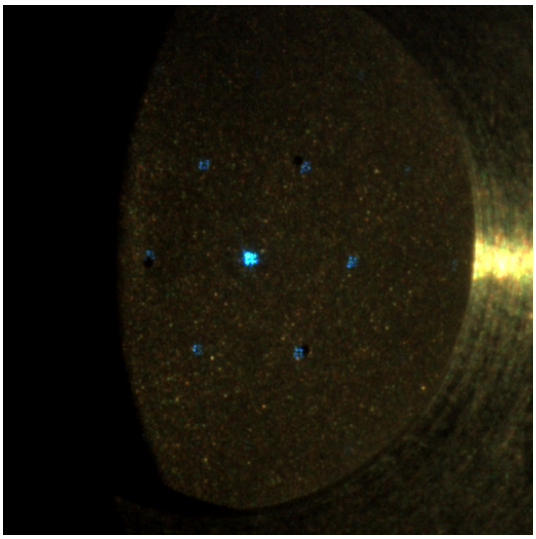


Figure 6-2. Aligning the diffracted orders with the pinhole mask. The image is captured by the XIMEA camera as an LED is illuminating the mask plane. In this image, the current positions of the diffracted beams require adjustment to align with the pinholes.

6.1.2. A Python based platform: HexSIM ScopeFoundry

Given the range of optical and mechanical hardware in the system, it is convenient to integrate their control and operation into one software package. As the hologram calculation and phase recovery algorithms run in Python, and Python packages are provided by manufactures of many devices, ScopeFoundry (a Python based platform) was chosen as a favourable platform to integrate all hardware controls and SIM calculations. Based on the basic frame in ScopeFoundry, a frame called 'HexSIM' was originally written by Hai Gong to control equipment in our system and visualise measurements in experiments. The HexSIM ScopeFoundry has then been further developed and modified as described here to execute imaging with updated devices such as the new SLM.

As illustrated in the block diagram in Figure 6-3, HexSIM ScopeFoundry integrates control of: the SLM for displaying gratings (holograms), the DAQ box for synchronisation in both the phase recovery and SIM imaging, the two lasers, the z stage and two cameras. A bespoke analyser called HexSIM Analysis for process raw images and analyse reconstructed frames is not included, which is not commonly used since the newer napari-SIM-processor, incorporating and improving on the original HexSIM Analysis, has made HexSIM Analysis redundant.

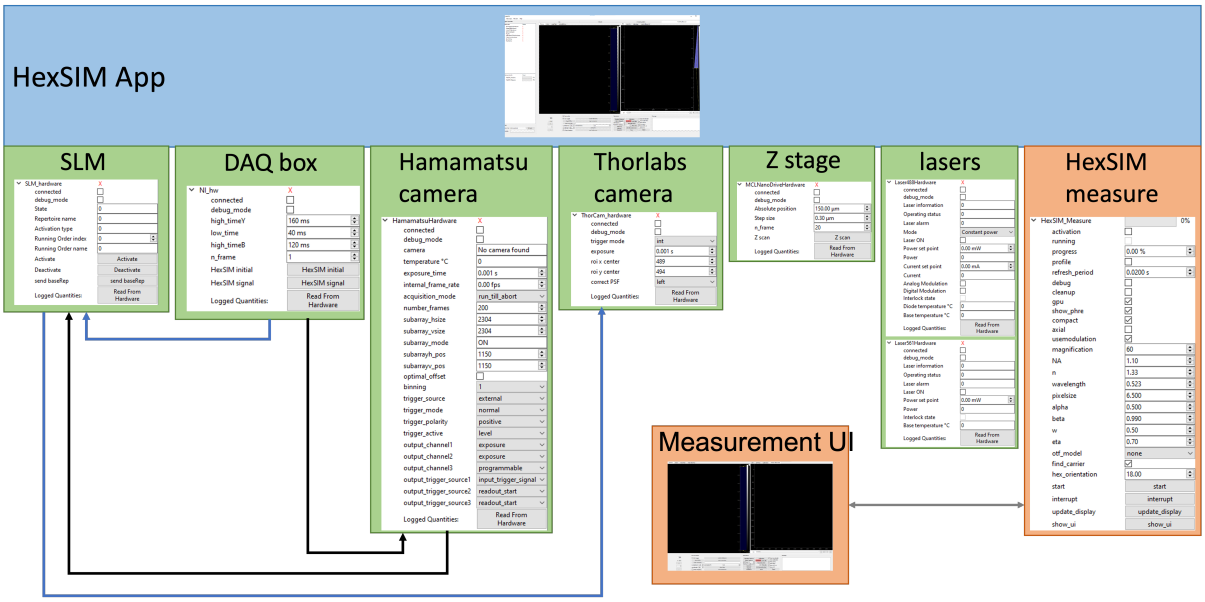


Figure 6-3. The block diagram of all the different device driver elements in our HexSIM system.

Figure 6-4 presents the graphical user interface (GUI) of the HexSIM ScopeFoundry main window. As can be seen in the figure, some essential devices are visible in the GUI as

hardware widgets, and measurements and analysations are customised as modules, with imaging and reconstruction of HexSIM visualised in windows. All widgets, modules, user interface controls are designed and tailored to the implementation of HexSIM imaging with phase measurement and recovery.

Figure 6-4. The graphical user interface (GUI) of HexSIM ScopeFoundry, including pieces of hardware, measurements, windows, functional buttons and so on. All tabs are collapsed in the screenshot, but can be expanded to control particular devices or measurements.

By using the ScopeFoundry framework, the level of user control for each device can be customised. For example, Figure 6-5 presents the user adjustable tabs for the Thorlabs camera used to observe the diffraction orders on the mask. As the camera is only used for alignment, not imaging, some camera control parameters are not critical and are therefore not presented to the end user, leaving a concise list of controls. All user interface elements are written in Qt for Python [100], which is the same toolset as used in napari.

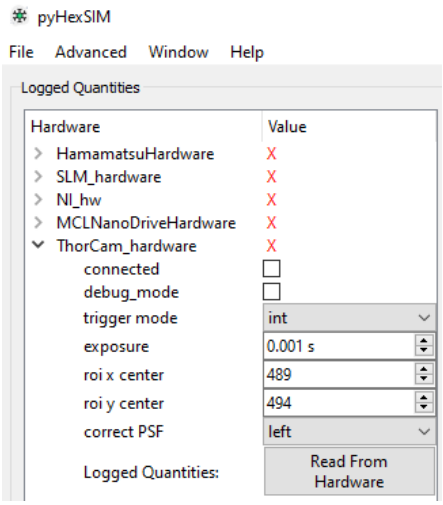


Figure 6-5. The control tab of the Thorlabs camera in the GUI of HexSIM ScopeFoundry.

In addition, the functions of each device can be added or removed from the ScopeFoundry GUI conveniently as required. We have found that the ScopeFoundry framework is a powerful tool for the control and integration of various devices in microscopy experiments.

6.2. Implementations in HexSIM ScopeFoundry

By using HexSIM ScopeFoundry, SIM imaging with hexagonal patterns can be implemented. Before performing SIM imaging sessions, the phase recovery can be implemented in-situ to correct the aberration caused by the non-flat SLM, and this procedure is only required to be repeated when the equipment such as the spatial mask and the alignment in the system are changed. SIM imaging is then conducted in HexSIM ScopeFoundry, with the SLM and the camera's exposure being synchronised by hardware triggering. Because the platform is customised according to our HexSIM and integrates several of pieces of equipment, it is not possible to elaborate all functions and designs here. However, the implementation of the in-situ phase recovery in the HexSIM system, the synchronisation between the SLM and the camera and the design of a repertoire to integrate the SLM's various functions and are introduced in this section.

6.2.1. In-situ phase recovery and applying phase stepping

The algorithms of correcting the phase aberration based on PSF measurements are explained in Chapter 4, and simulations using different fitting functions are demonstrated. In the in-situ scenario, the algorithm with the sampled Chebyshev polynomials is used to measure and recover the phase. The direct hologram transfer and the hardware triggering between the SLM displaying PSFs and the imaging using the Thorlabs camera are the same as in the case of the simple optical layout (see Chapter 4). The difference between implementations in the simple layout and the HexSIM scenario is that, as demonstrated above in Figure 6-1, the beam sampler is installed to partially direct the incoming light from the SLM for the in-situ PSF measurements. The light source used for in-situ correction is the blue laser (Oxxius, LBX-488) as its power can be adjusted to a low value to avoid over exposure in PSF measurements. With the low power laser, the PSFs are imaged by the Thorlabs camera, and the iterative algorithm of the phase recovery is executed. In HexSIM ScopeFundry, the algorithm for correcting the phase aberration is integrated with hardware control of the equipment such as the SLM, the Thorlabs camera and the laser for PSF measurements.

The result of the phase correction is used in the DBBI method to design holograms for HexSIM. Although the measurement is made using the blue laser, the resulting correction is easily applied to the yellow laser. The holograms also depend on the orientation of the pinholes on the mask and the distances between them. Using a tailored GUI element in HexSIM ScopeFoundry, these physical characteristics of the system can be conveniently added into

the hologram design. Especially when the mask is changed, it is easy to correspondingly adjust the hologram calculation.

6.2.2. Synchronisation between the camera and SLM: hardware triggering

In addition to the phase correction, the holograms can also include phase stepping. For our two-colour system using hexagonal illumination patterns, fourteen stacks of holograms are generated, seven for each laser illumination wavelength, with each stack consisting of hexagonal patterns with different initial phase steps. In the simulations presented in Chapter 5, all resulting illumination patterns as the initial phase is stepped are summed up, while in the practical implementation, all holograms with different phase steps are displayed sequentially at a high speed and the resulting images are integrated during the relatively long exposure of the Hamamatsu camera. In this way, the uniformity of the phase and amplitude of frequency carriers can be improved. Additionally, the possibility of permanently photo-bleaching the illumination patterns into the sample is reduced since the structured illumination is always changing during the imaging of a set of SIM raw images. Managing the display of all these holograms and timing the associated acquisitions on the camera are described in this section.

Even though all the hardware can be integrated and controlled directly using HexSIM ScopeFoundry from software, the imaging process would be unduly time consuming if control buttons for the SLM and the Hamamatsu camera requires to be pushed successively. In particular, the synchronisation between the integrated hardware devices is essential for the imaging in the HexSIM system, and it is desirable for devices to be hardware triggered to achieve a high imaging speed. To achieve the high speed of imaging, the SLM that displays holograms and the camera which captures raw images are synchronised by hardware triggering as shown below in Figure 6-6 and Figure 6-7.

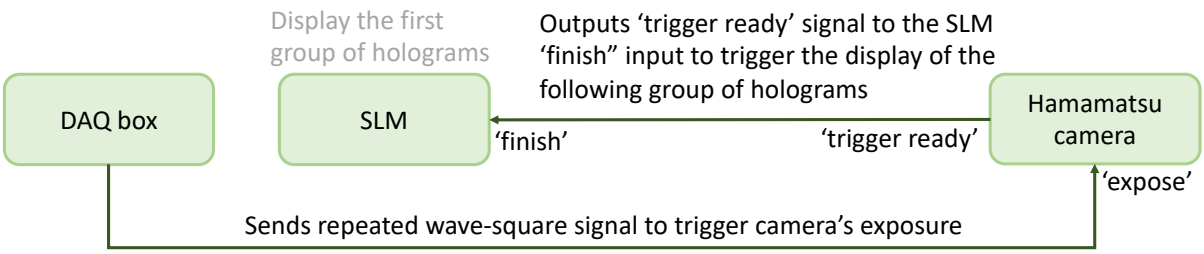


Figure 6-6. The block diagram of the configuration of synchronising different device for SIM imaging. The initial step is written in a light grey shade.

A data acquisition device (national instruments, NI USB 6351) is used to generate signal of repeated square waves to hardware trigger the camera's exposure. As presented in Figure 6-7, HexSIM imaging starts with the SLM displaying the first group of phase stepped holograms, corresponding to the first raw image in the green channel. The DAQ box is used to produce a signal consisting of successive pulses defining the required exposure for the two-channel in our HexSIM system. The trigger mode of the Hamamatsu camera is set to external level trigger mode that integrates the intensity captured over a period corresponding to the high levels in the triggering signal. However, it does so using a rolling shutter, such that the later rows in the camera to be read out are exposed for the same exposure period but starting a time after the first rows corresponding to how long it takes to read out all the intervening rows. During this whole rolling shutter period the SLM hologram should remain consistent. Then, as the readout of each image finishes, the loop displaying the current group of holograms in the SLM driver is set to stop on receiving an external 'finish' signal from the camera. One of the digital outputs on the camera is set to give a rising edge 'trigger ready' signal at the end of readout that is connected to one of the SLM driver's input digital pins as this 'finish' signal. Thus, when the readout ends, the SLM is triggered by the camera's output to start displaying the following group of holograms.

As presented in Figure 6-7, when capturing each raw SIM frame, ten holograms with different initial phase steps are uploaded on the screen of the SLM in sequence, which are periodically displayed as a loop. The duration of presenting each hologram is relatively short (e.g. 2 ms) compared with the exposure time of the camera frame as a whole (e.g. 120 ms). The high periods of the green and the red channel pulses from the DAQ-box can be adjusted according to the power setting of the lasers and the brightness of the sample but is set to multiples of the complete phase-shifted hologram set (e.g. 20 ms). The low time between red and green exposure pulses is set to the same value but should be at least as long as the read-out since this length is related to the size of the captured images, which is the same for all the images in the two channels. The exposure of the camera stars at the first line in the sensor with the rising edge of the external 'expose' trigger signal. Because both the exposure and the readout of the camera uses a rolling shutter, one whole raw image is obtained only after the last line in the sensor is read out. The camera's exposure is shown as the light lilac areas in Figure 6-7. When the readout of the last line in the camera's senser ends, the camera outputs the 'trigger ready' signal, which is sent to the SLM to stop displaying the current group of holograms and change to the following group. The group won't change immediately since it needs to wait for the end of the current group. Hence, there needs to be some allowance made in the gap between exposure signals to include both the rolling shutter and the possibility that we might

have to wait 20ms for the current group to end. All holograms (either for the green- or red-channel imaging) with phase stepping are presented as the coloured line, which denotes repetition of the phase stepped stacks of holograms for each different frame in the SIM acquisition. In practice, the SLM requires to maintain DC balance by displaying a hologram and then its inverse, subsequently, which is not presented in the figure. As a result, when displaying a stack of 10 different phase steps, actually 20 different holograms are display with 10 of those being the inverse of the other 10. This is equivalent to a π radians shift in the initial phase step. With hardware triggering, the hologram presentation and acquisition of images using the camera are synchronised and can be achieved.

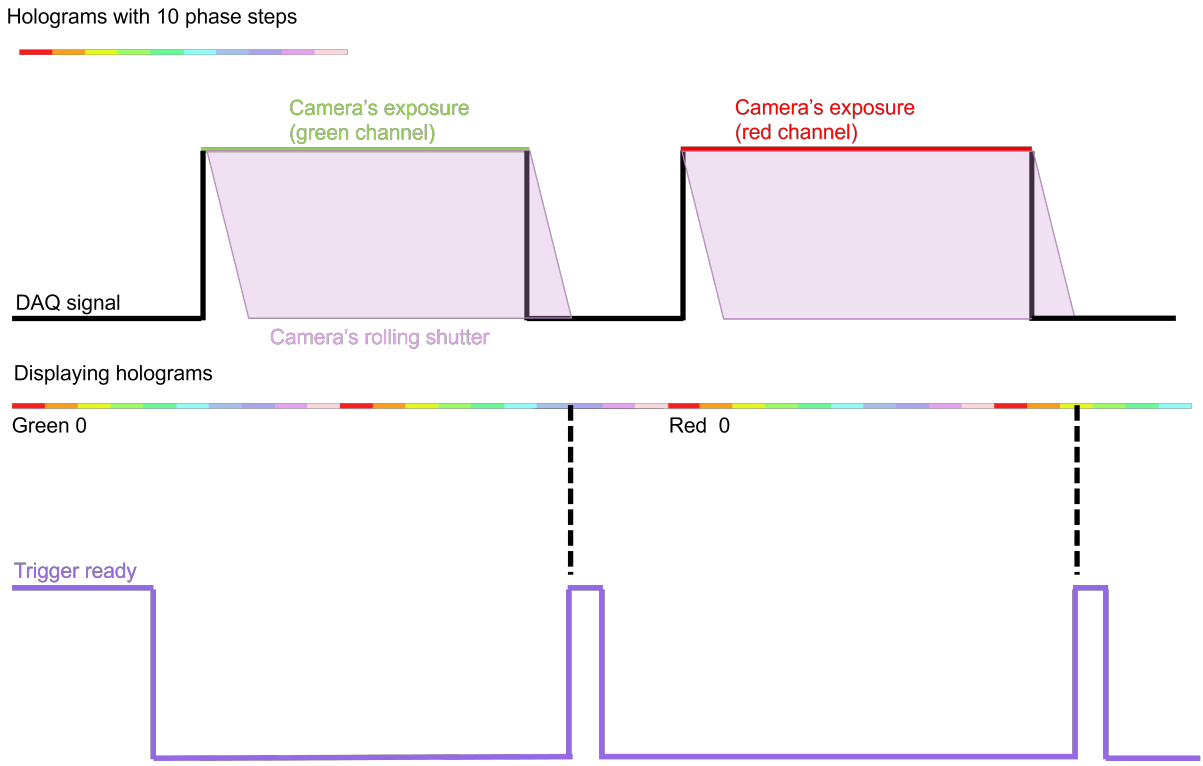


Figure 6-7. The synchronisation between the SLM and the Hamamatsu camera. The holograms for each frame at a channel (such as the Green 0 and the Red 0) are different, but they are shown using the same symbol of coloured lines. The light lilac areas demonstrate the exposure of the camera from the first line to the line at the bottom of the frame. The synchronised imaging for one frame of a raw image in two channels is presented in the figure, respectively. Displaying inverse holograms for the maintenance of the SLM's DC balance is omitted in the figure for clarity.

6.2.3. Basic repertoire loaded on the SLM

It is obvious that the SLM plays a crucial role in the HexSIM system. First, the SLM displays holograms for both the PSF measurements in the phase recovery and the generation of

illumination patterns with the phase stepping. Second, the output signal of the SLM can be utilised to hardware trigger the exposure of the Thorlabs camera in the case of correcting phase, and the SLM can be hardware triggered by the Hamamatsu's 'trigger ready' to change the loop of displayed holograms when SIM imaging. Third, the SLM can display updated holograms continuously to facilitate alignment of the diffraction orders and the pinhole mask remotely. To carry out these various functions, a basic script (which is called a repertoire) for the SLM is designed and sent to the driver board of the SLM.

In general, the repertoire includes information such as the duration for displaying each hologram, the different mode of the activation of the display, the sequences of images to display as holograms, and so on. The instructions of controlling and determining the display of a specific set of holograms is called a running order, and there can be multiple running orders in one repertoire. In our implementation, a basic repertoire including various running orders is composed and sent to the SLM board from the running HexSIM ScopeFoundry instance. However, the holograms are not included in this basic repertoire since they might need to be modified during experiments. The basic repertoire only arranges the locations of holograms in various use cases, while the holograms can be updated by directly writing images onto the SLM's flash. To conduct different scenarios in the experiment, as demonstrated in Table 6-1, the bit-planes where holograms reside at are assigned in the flash of the SLM, for the various running orders (ROs). Since each bit-plane on the flash memory cannot be rewritten unlimitedly, there are multiple ROs are assigned for each use case that are linked to different areas of the flash buffer. In the experiment, when the SLM is used in a particular scenario, a RO can be randomly selected from the appropriate set the repertoire by HexSIM ScopeFoundry to make sure that the same set of bit-planes in the flash will not be updated repeatedly. As shown in the last column in the table, the arrangement covers all bit-planes in the flash (768 frames), except for the first nine bit-planes which are reserved for other functions of the SLM.

Scenarios	RO repeats	Bit-planes per RO	RO number	Bit-planes used
Reserved bit-planes				0 - 8
Checking PSFs without hardware triggering	7	1	0 - 6	9 - 15

Checking 3 beams without hardware triggering	3	21	7 - 9	16 - 78
Displaying hexagonal holograms	4	140	10 - 12	79 - 638
Phase correction	43	3	14 - 56	639 - 767

Table 6-1. Assignments of bit-planes in the flash memory of the SLM and the corresponding running orders. Each scenario describes a function which the SLM can execute.

6.3. Results of in-situ phase recovery in the HexSIM system

Given the methods explained in Chapter 4 and the implementation introduced above, the phase aberration caused by the curved SLM can be measured and corrected. As demonstrated in the schematic (Figure 6-1), the blue laser (Oxxius, LBX-488) is used with a power set to 2 mW to avoid the overexposure of imaged PSFs. The laser source is expanded before illuminating the SLM's surface where a sinusoidal hologram is displayed. The source light is then diffracted, and two light orders are focused by the lens 1. One of two orders is imaged by the camera 2 at the centre of a selected region-of-interest field of view, and the acquired PSF images are used in the phase-recovery method to indirectly measure and correct the phase of the SLM. The criterion is that the RMS must be less than 0.15 radians in two consecutive measurements.

By applying the combination of the Gerchberg-Saxton algorithm, PSF measurements and the fitting to Chebyshev polynomials, the phase of the SLM can be retrieved. Figure 6-8 shows the progression of the in-situ measuring and correcting the SLM's phase, including the PSF measurements, the calculated PSF, the estimated pupil phase and intensity, the unwrapped and Chebyshev fitted pupil phase and the accumulated phase estimate.

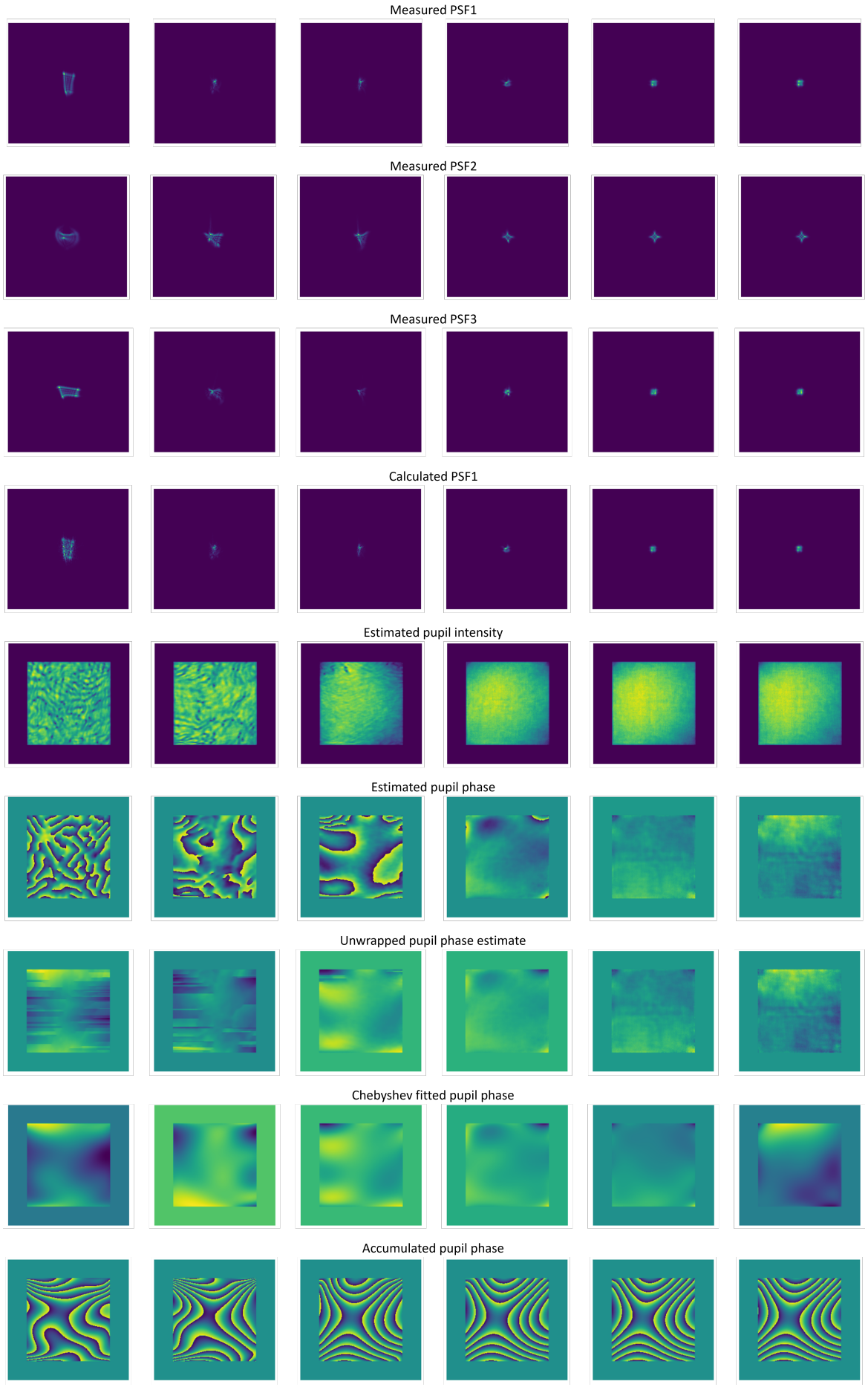

Measured PSF1
Measured PSF2
Measured PSF3
Calculated PSF1
Estimated pupil intensity
Estimated pupil phase
Unwrapped pupil phase estimate
Chebyshev fitted pupil phase
Accumulated pupil phase

Figure 6-8. The progression of the in-situ measurement and the recovery the phase of the SLM on the HexSIM layout. The phase recovery process is depicted from the left column to the right.

Figure 6-9 shows a typical graphical output as the aberrated phase is retrieved at the end of the phase retrieval process. The measured diffraction orders (PSFs) with the current correction and artificially added bias aberrations are shown in (a) to (c), and the matching PSF generated by the G-S process for that in (a) is shown in (d). The estimated residual phase of the SLM before and after the unwrapping is shown in (e) and (h), respectively. As shown in (f), the estimated intensity of the pupil is nearly uniform. The values of the Chebyshev modes for the current residual phase are plotted in (g). Compared with the noisy image of the retrieved phase in (h), the fitted one shown in (i) is much smoother. As recorded in (j), this phase correction finishes in five iterations with a residual RMS phase of 0.113 radians. The RMS is markedly smaller than the diffraction limited value of 0.469 radians. The accumulated phase and one example of the hologram written into the flash of the SLM are illustrated in (k) and (l), respectively. The entire process finishes in only about 26.15s.

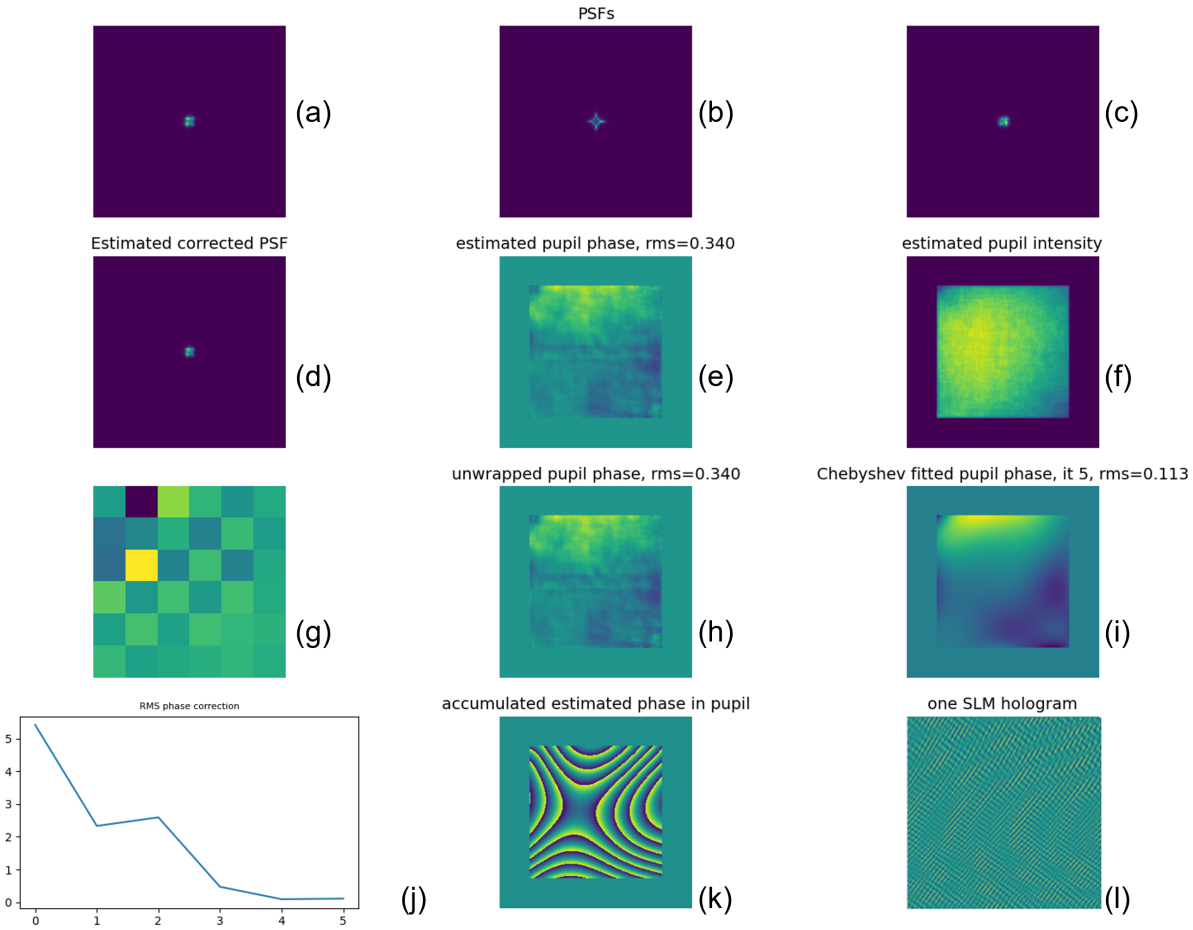


Figure 6-9. The in-situ measurement and the recovery the phase of the SLM on the HexSIM layout. (a) – (c) Three corrected PSFs with introduced biases. (d) The calculated PSF with the 1st bias. (e) The calculated pupil phase with rms as 0.340 radians. (f) The calculated pupil intensity. (g) The padding of values of Chebyshev modes, and the square on the top left shows the piston value. (h) The estimated unwrapped phase with rms as 0.340 rad. (i) The Chebyshev fitted phase with the 0.113 rms. (j) The rms of the fitted pupil phase. (k) The estimated phase accumulation. (l) The hologram written into the flash of the SLM to diffract the order in (a). All results shown are obtained from the last correction iteration.

Subsequantly, as shown in Figure 6-10 (b), the three diffraction orders pass through the pinholes completely, while the corresponding orders shown in (a) are partially blocked by the pinholes. Clearly, the algorithm has successfully correct the aberrated beams to diffraction limited ones. Two beams on the bottom right are framed by white rectangulars as examples to compare the situation before and after phase recovery. The images are remotely obtained using the Ximea camera while illuminating the pinhole mask with a white LED.

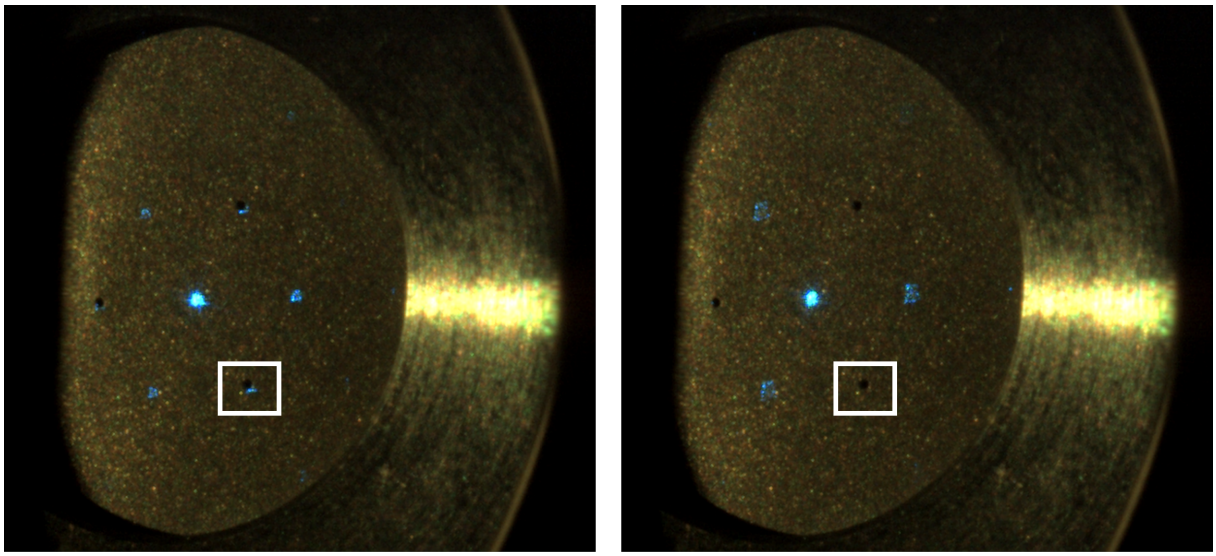
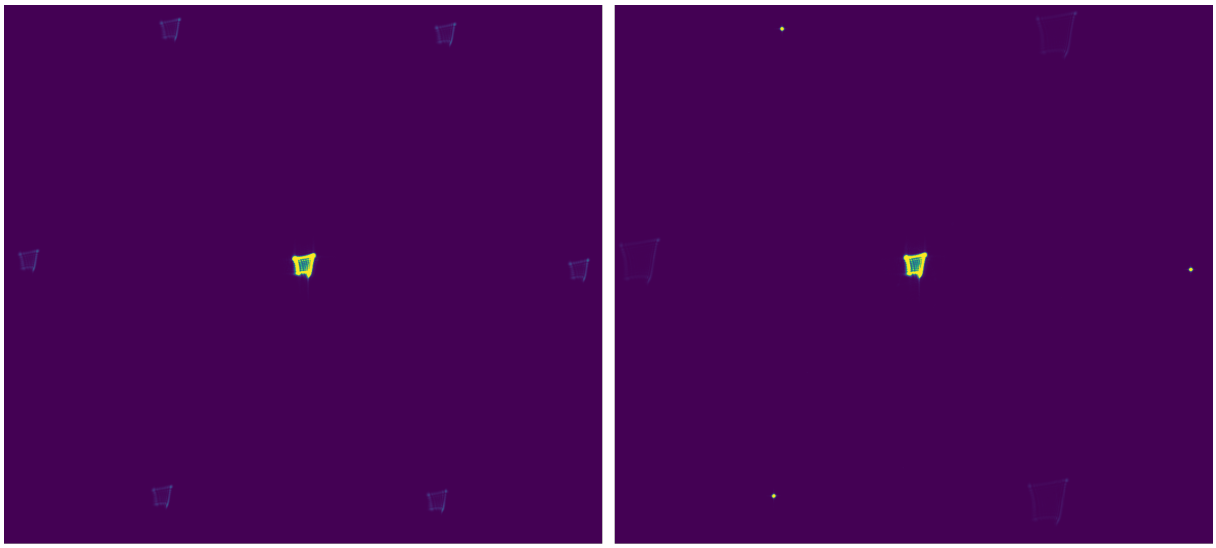


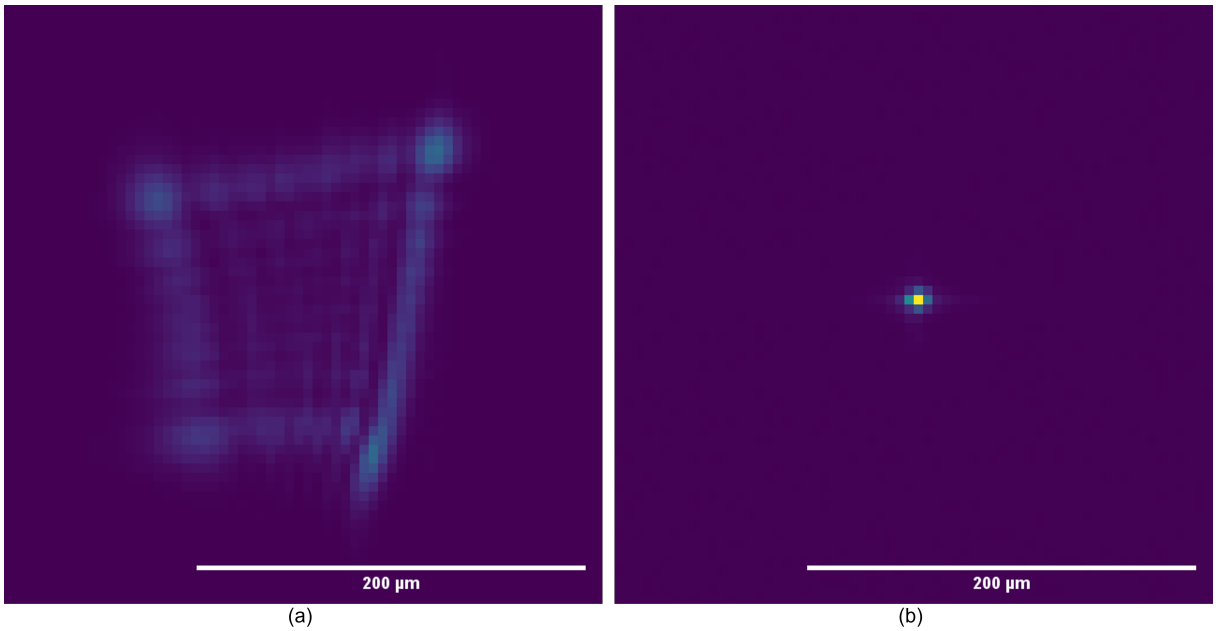
Figure 6-10. The diffraction beams locating at the mask before (a) and after (b) applying phase recovery to correct for SLM flatness. The two white rectanguls are used to denote the diffracion orders passing through the pinholes in two cases. Images are taken with camera 3.



(a) (b)

Figure 6-11. Diffraction orders captured by the Thorlabs camera. (a) Uncorrected diffraction orders. (b) Corrected diffraction beams. The extended and weak three diffraction orders are conjugate to thoses are corrected. Images are over exposed to display weak orders.

Images of the diffraction pattern obtained using the Thorlabs camera can also show the effectiveness of the phase recovery and correction, as demonstrated in Figure 6-11. Comparing the diffraction orders before (a) and after (b) the phase correction, it it obvious that three beams are corrected, while the other three conjugates now see double the aberration, being more extended than the uncorrrected case. (a) and (b) are over exposed so that the extended diffracted orders are visible. The comparison of one aberrated diffracted order and the corresponding corrected one are shown in Figure 6-12 (a) and (b), respectively. By applying the in-situ phase-recovery method, in contrast to the expanded diffracted order (a), the corrected one (b) is small enough to reach the limit of the detector. The detected PSF in (b) is focused in just one pixel, which is limited by with the pixel size of the camera as 4.8 μm. Although the real size of the PSF could be smaller. The 200 μm scale bar in the images denotes the diameter of the pinhole on the spatial mask.



(a) (b)

Figure 6-12. (a) One beam without the phase correction. (b) One recovered diffraction order. The diameter of the pinhole on the mask is 200 μm.

This method is very effective for phase measurement and correction, and only needs to implemented once before carrying out SIM imaging, for as long as the rest of the experimental set-up remains unchanged.

6.4. Measurements on fluorescent samples

After the implementation of the phase measurement and correction, SIM imaging with hexagonal illumination patterns can be performed on fluorescent samples. The results with and without the phase recovery are compared, and the improvements attributed to holograms applying the phase stepping are discussed. Measurements are carried out on BPAE (bovine pulmonary artery endothelial) cells (ThermoFisher Scientific, FluoCells #1), with Alexa Fluor™ 488 phalloidin labelling the F-actin cytoskeleton (green channel) and MitoTracker™ Red CMXRos used to stain the mitochondria (red channel). Various hexagonal holograms generated by the DBBI method with or without the phase correction. Using two illumination lasers with different wavelengths two channels are imaged, processed and then merged into red/green images when shown below. All raw images are processed using the plugin napari-SIM-processor. All the images presented below are with a mask pinhole of diameter 100 μm placed on a 2.4 mm radius circle.

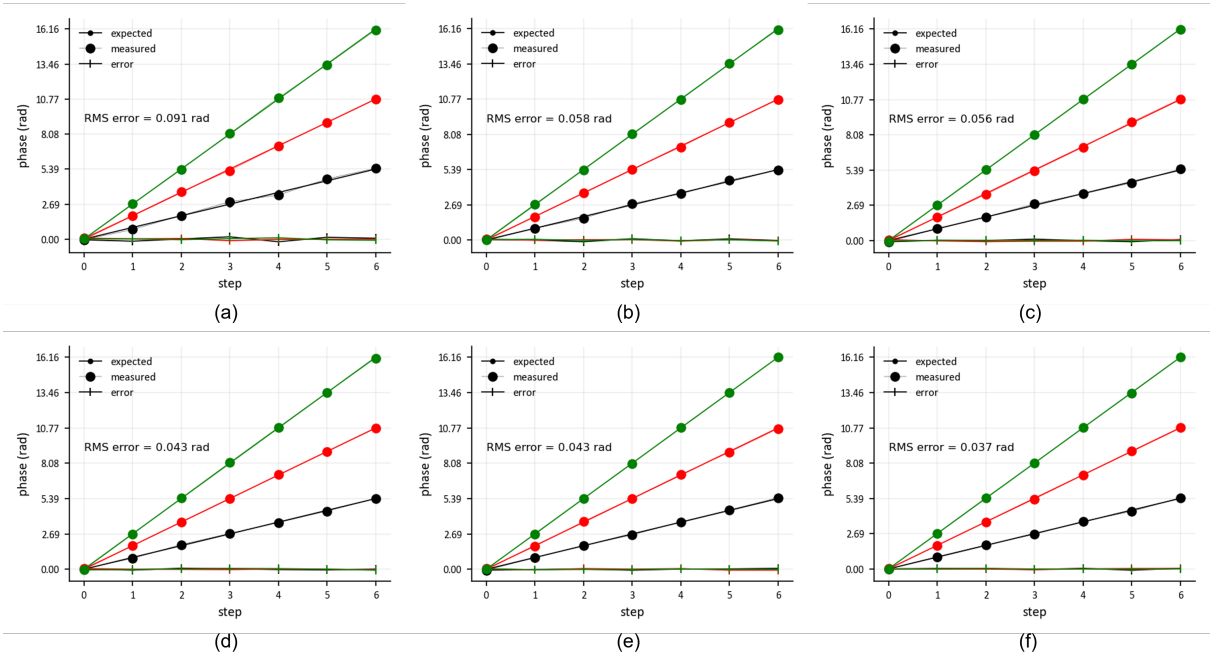


Figure 6-13. Calibrated phase versus hologram step. Green channel (a), (b) and (c) and red channel (d), (e) and (f). Without correction (a) and (d), with correction (b) and (e), with correction and phase stepping (c) and (f). DBBI method used for all hologram designs. The RMS error is shown in each sub-image.

The DBBI method is used to generate holograms both with and without phase correction and phase stepping, and the obtained raw frames are processed using napari-sim-processor (introduced in Chapter 3). The first step in processing the raw data is to measure the carrier frequencies, phases and amplitudes and the resulting plots for phase shown in Figure 6-13 for both green and red channels and for the three scenarios of no-correction, correction and correction with phase stepping. The phase of each of the three carriers is expected to increase by $\frac{2\pi}{7}$, $\frac{4\pi}{7}$ or $\frac{6\pi}{7}$, respectively, at each hologram step. Clearly, and in all cases the holograms used on the SLM have generated carriers in the sample with close to the correct expected phase step. Some discrepency is expected as these calibrated phases are based on the measurements of images from the real sample and their accuracy is somewhat dependent on the information content in the sample to be able to correctly identify the carriers in the illumination. Perhaps in this respect Figure 6-13 (a), without any correction for the SLM flatness, does not show such a close agreement (the RMS error is 0.091 radians), reflecting poorer raw image quality. In the green channel, RMS error decreases to 0.058 rad due to the phase correction (see b), and declines further to 0.056 rad when the phase stepping is added (see c). For the red chennel, the phase recovery and phase steeping case (f) shows the least RMS error as 0.037 rad, and the values in other two cases (d) and (e) are the same as 0.043 rad since the signal in the invisible FOV in the case without any correction is not significant. In

all situations, the DBBI method with phase correction and phase stepping provides the minumum RMS error in the phase determination. These calibration measurements are also based on the whole image so do not highlight other local variations in phase or amplitude that might be affected by the phase-stepping correction also applied in (c) and (f). Overall these plots validate the hologram design processes used.

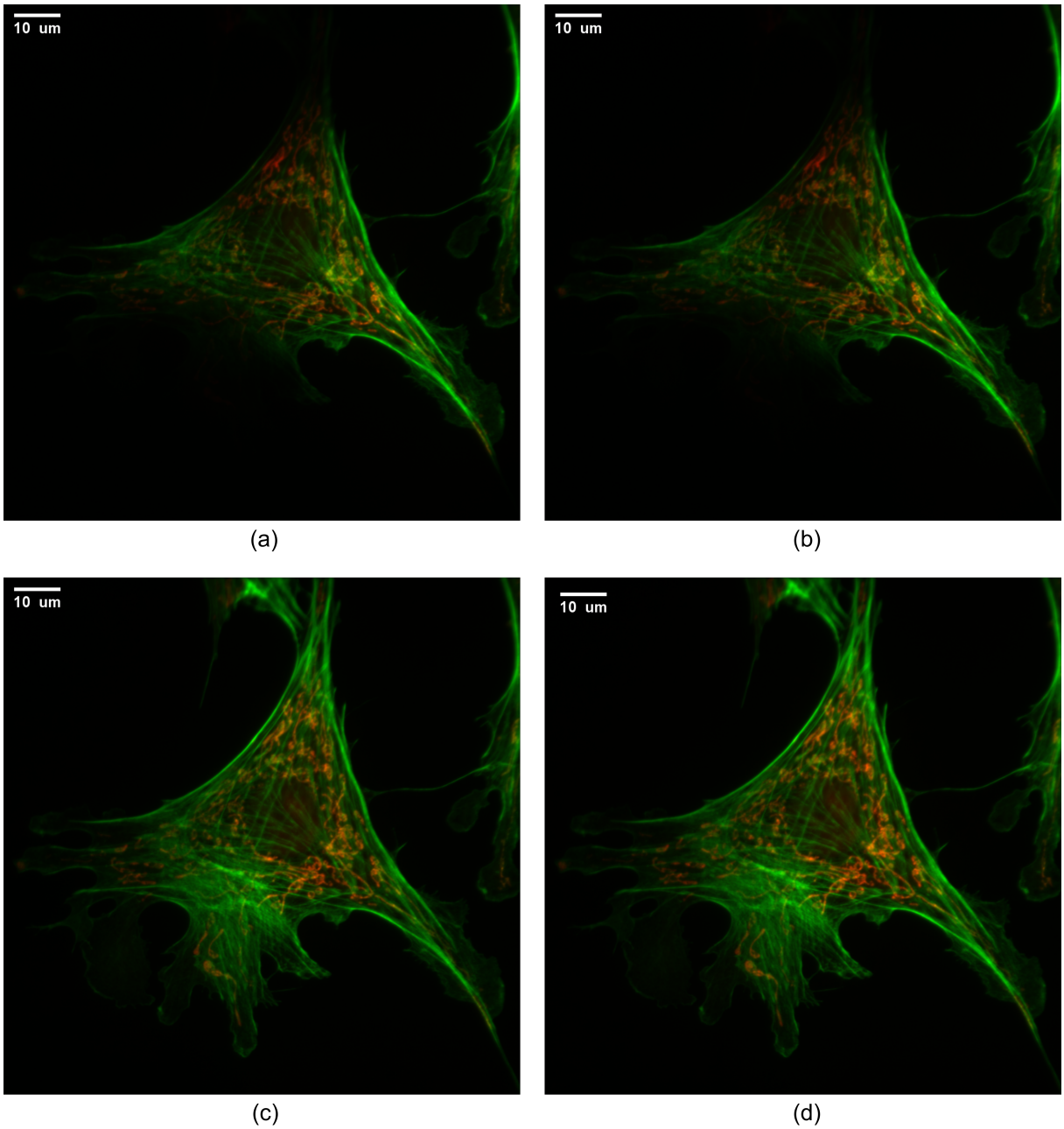


Figure 6-14. Wide-field images for four different cases of hologram generation with different modulation strategies. (a) The wide-field image with holograms generated by the DBBI method. (b) The DBBI with phase stepping is utilised to generate holograms. (c) Holograms are produced using the DBBi method with phase correction. (d) The wide-field image with the phase correction and phase stepping. All holograms used to

construct illumination patterns are generated using the the DBBI approach. The four wide-field images are produced using napari-sim-processor as the zero-order carrier image of the acquired stack.

Figure 6-14 shows the wide-field images that are calculated by the processing algorithm based on the DC carrier component. As shown in the top row in Figure 6-14 (a) and (b), the FOV of the wide-field image is apparently incomplete compared to those where the DBBI method includes phase correction, (c) and (d). Without correction the wide-field images show only partial illumination in the the top, left and bottom parts of the cell compared to images with correction. The image with both phase correction and phase stepping is presented in (d), which has slightly more uniform brightness than that without phase stepping shown in (c). For the holograms including the phase correction used in (c) and (d), the three diffracted orders can all pass through the pinholes, providing a complete field of view (FOV) in the raw images, thus in the wide-field images. Because of the large FOVs, the results with the phase correction can present more complete information about the sample than those cases where three diffracted beams are partially blocked. Because the wide-field image obtained with holograms that only includes phase stepping is equally affected by the aberrated phase as that without, neither displaying the entire FOV (b), this approach is not discussed further below.

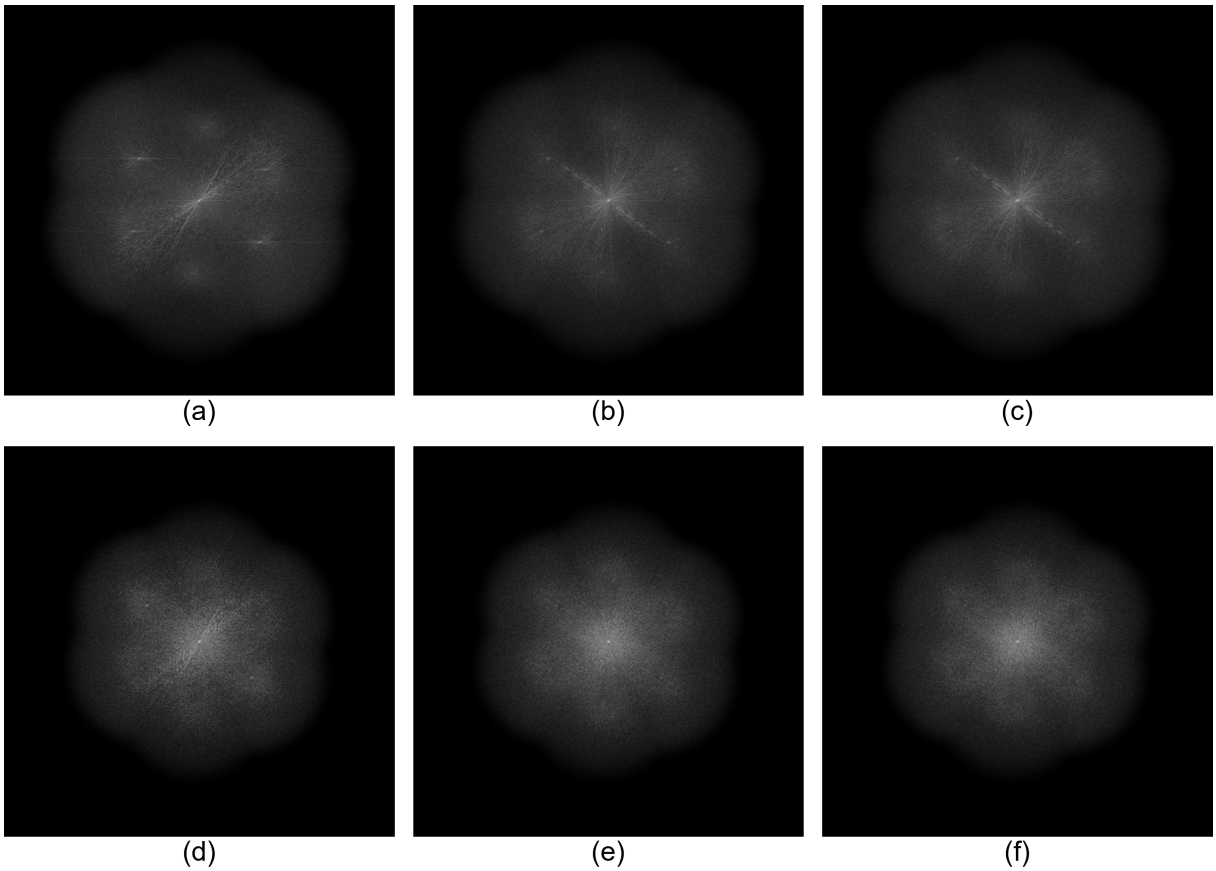


Figure 6-15. The spectrum of reconstructed images in green channel without phase correction(a), with phase correction (b) and with both phase correction and phase stepping (c). (d), (e) and (f) shows the corresponding spectrum of the reconstructed images from the red channel.

The raw images in three scenarios (without correction is used, with correction, and with correction and phase stepping) are reconstructed. An analysis of the reconstructed data in frequency space is presented in Figure 6-15. Here again the scenarios without SLM flatness correction (a) and (c) yield significantly different frequency spectra to those with (b), (d), (e) and (f). In the green channel in (a) whole spatial frequencies are not measured or reconstructed that can be seen in (b) and (c), while there are strong peaks at the carrier frequencies present that will show up as patterned artefacts. As presented in (b) and (e), there appears to be some extra signal near the carrier frequencies in the spectra without phase stepping, which might be expected as this would be the artefacts resulting from interference with higher diffracted orders in adjacent pinholes. In the red channel the differences in spectra between uncorrected and corrected datasets is not so stark yet still clear differences exist. There is more frequency noise in (d) compared to (e), while (f) has the least amount of noise, which is consistent with our expectation. In the green channel (top row), which is used for imaging filaments in cells, the spectra still appear non-ideal despite aberration correction and phase stepping. This is due to the filaments' inherent directional properties, which cause noticeable lines in the frequency images.

The reconstructed images are shown in Figure 6-16 (a), (b) and (c), respectively. Compared to the reconstructed image presented in (a) that with a limited FOV, the images in the other two cases (b and c) are both of a complete FOV. Without correcting the phase aberration, the reconstructed result (a) has a lower contrast than those with the phase recovery (b and c), especially in the green channel. The yellow frame in (a) –(c) is shown in (a1), (b1) and (c1), respectively, showing the finer details of the reconstruction in green and red channels. There are some honeycomb artefacts in the filaments presented in (a1), while the green-channel information in (b1) and (c1) is resolved to show the clear filaments. The red-channel reconstruction shown in (a1) is apparently dimmer than those in (b1) and (c1). Comparing the scenarios of the DBBI with only phase correction (b1) and the method with both phase correction and stepping (c), the reconstructed images are similar, although details are different and there are more brighter pixels presented in (b1), whose values are higher than other pixels nearby, which may be due to residual pattern artefacts. It appears that these artefacts are reduced by the introduction of phase stepping into the DBBI method.

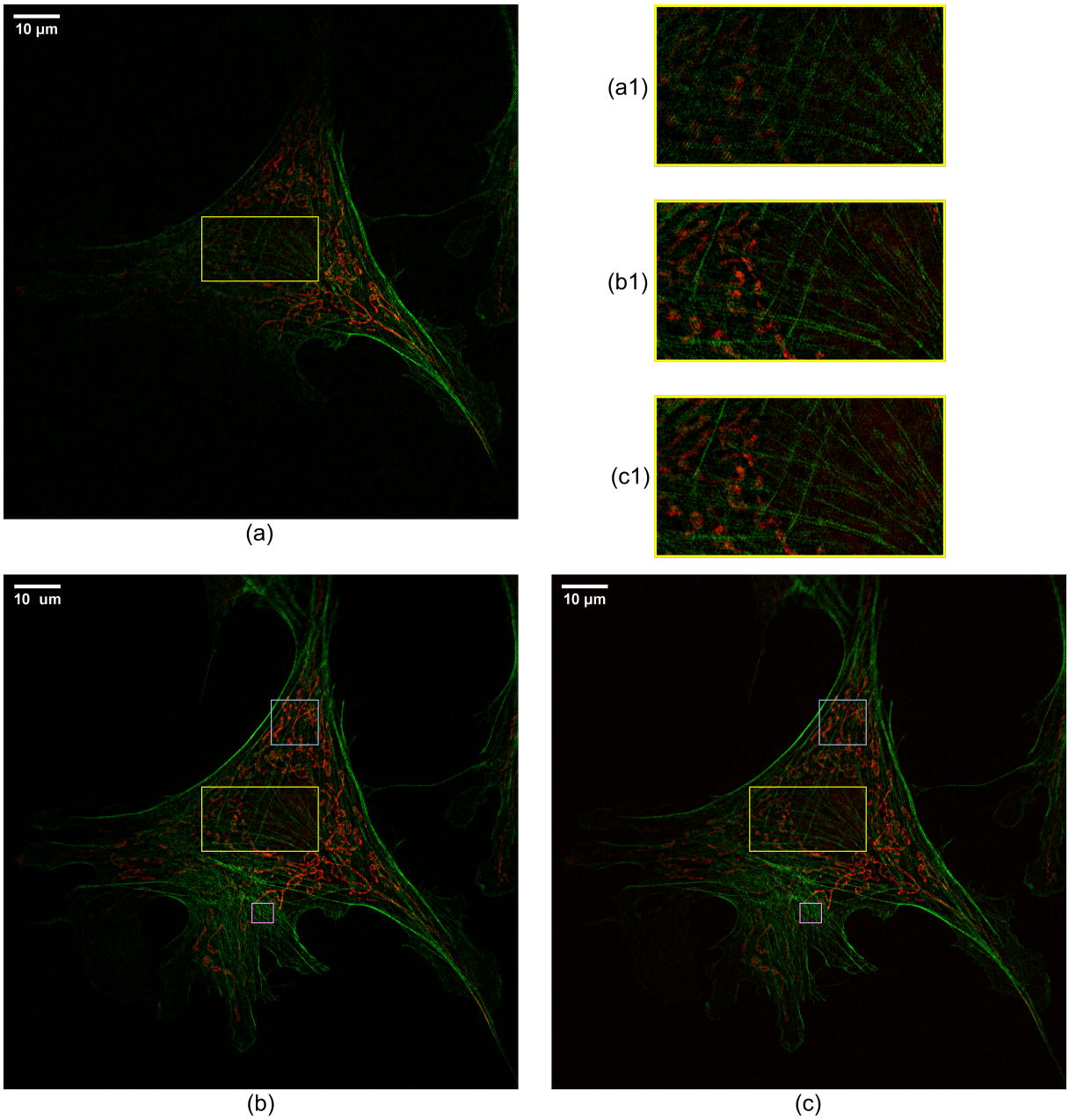


Figure 6-16. Reconstructed images using the DBBI method in different situations. (a) The case using DBBI without any modulations. (b) Utalising DBBI with the phase recovery. (c) The DBBI method is exploited with phase correction and phase stepping. (a1) – (c1): The yellow frame in (a) – (c), respectively. All imgaes are reconstructed using napari-SIM-processor.

To further compare the scenarios without and with phase stepping when the aberrated phase is recovered using the DBBI method, the details framed in blue and pink shown in Figure 6-16 (b) and (c) are presented in Figure 6-17. As shown in the red channel in Figure 6-17 (b2), there are some higher-intensity pixels, which seem to break up linear structures or filaments, while the brightness in (c2) is more uniform and the structures are smoother. This means that the DBBI combining correcting the phase aberration and phase stepping can smooth the

brightness in reconstructed results. As presented in the green channel in (b3) and (c3), the number of bright pixels in (b3) is obviously higher than that in (c3).

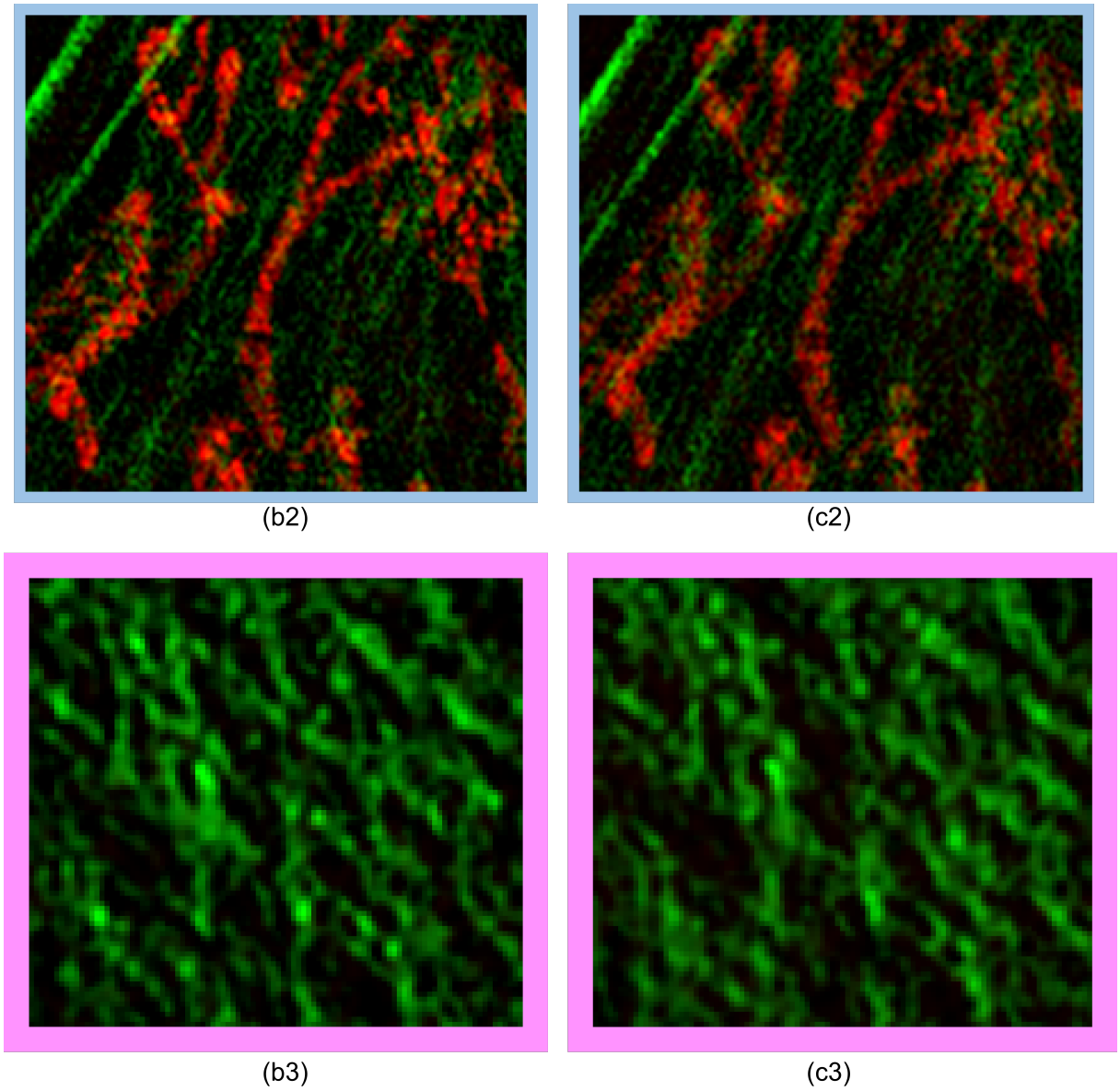
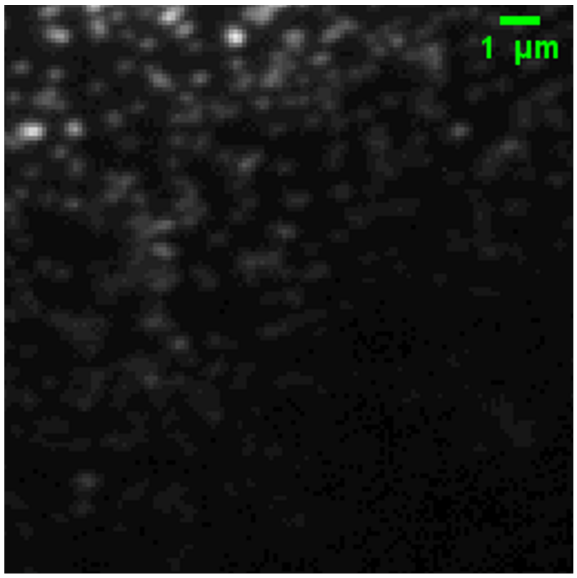
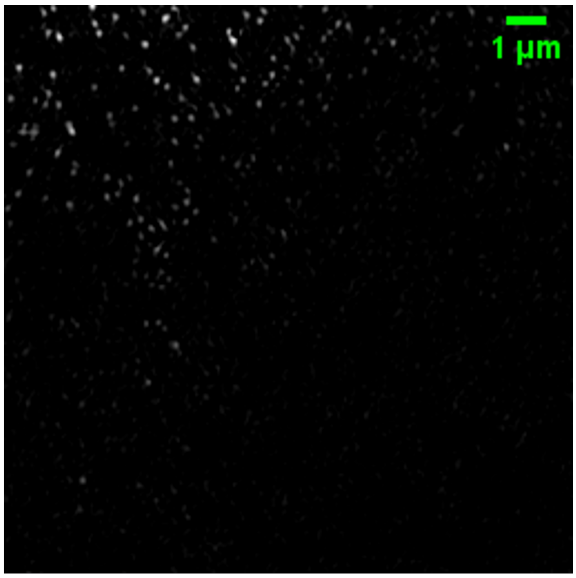


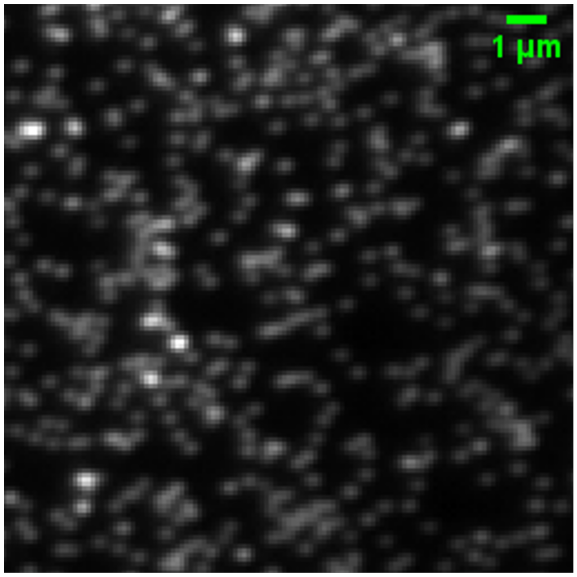
Figure 6-17. Comparison of reconstructured HexSIM images. The holograms are produced using the DBBI method without (left column) and with phase stepping (right column), where the phase aberration is corrected. (b2) and (c2): The red-channel information with and without phase stepping, respectively. (b3) and (c3): The green-channel image exluding phase steeping and including phase stepping, respectively. The blue and pink frames correspond to those in Figure 6-16 (b) and (c).



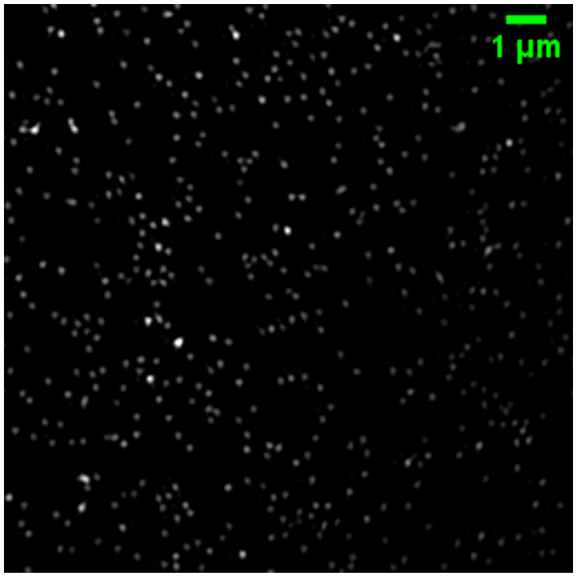
(a)



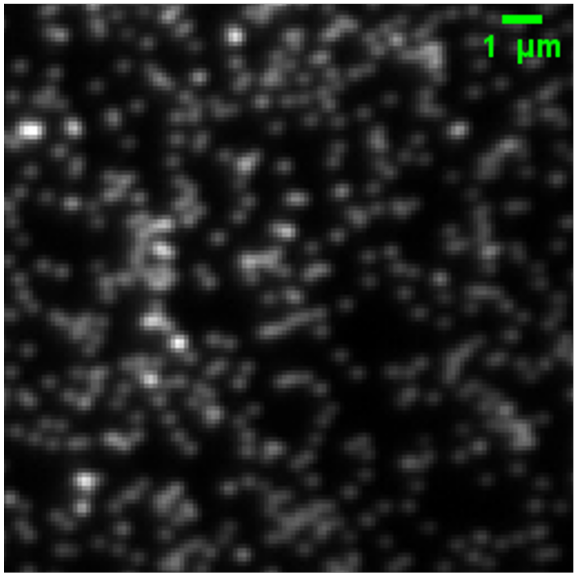
(b)



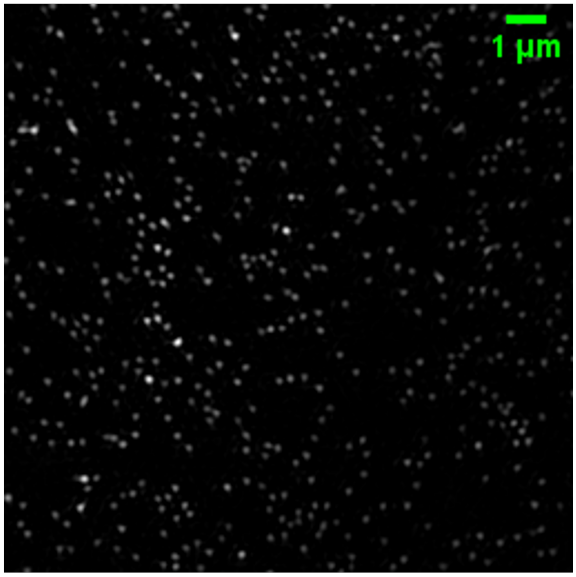
(c)



(d)



(e)



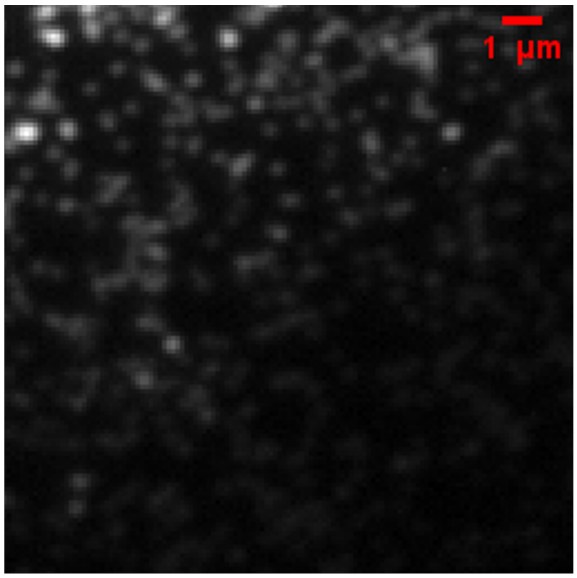
(f)

Figure 6-18. Wide-field images (a), (c), (e) and reconstructed images (b), (d), (f) of 100 nm beads in the green channel. (The excitation laser features 488 nm. The distance between a pinhole to the centre of the spatial mask is 2.3 mm.) The first row shows images without any corrections, the second row shows results with phase recovery, and the third row shows results with both phase recovery and phase stepping. All holograms are generated using the DBBI method.

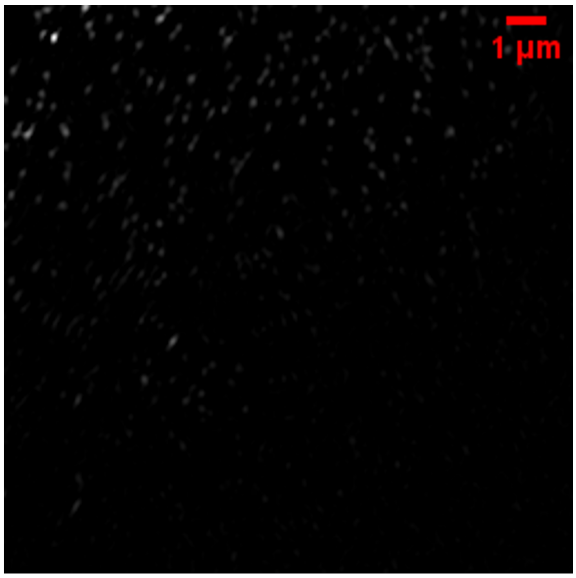
However, it is challenging to compare the results between phase recovery alone and phase recovery combined with phase stepping. In order to try to assess the difference between artefacts from the two different images, a ground truth is required. 100 nm beads (TetraSpeck™ Microspheres, 0.1 μm diameter) are imaged as a sample that is sub-diffraction limited with well-defined features that are relatively uniform and known. An image J plugin: ThunderSTORM [101] is usually used for single molecule localisation microscopy and works by fitting the diffraction limited images of single molecules to a gaussian mainly to find the location of the molecule. In this case it can give us an idea of the resolution of the SIM images based on the width of the fitted gaussians.

Obviously, as shown in Figure 6-18, when there is no correction in the hologram design, the FOVs of the wide-field image (a) and the resonctruction (b) are both not as complete as other sub-images. The shapes of beads in (a) and (b) are distorted. (c) and (d) shows the phase recovery results in the wide-field and the reconstructed case, respectively, showing complete FOVs. The DBBI method with phase recovery and phase stepping is used to generate the hologram, and the wide-field and reconstructed image is shown in (e) and (f), respectively. Compared with the wide-field images (c) and (e), the reconstruction (d) and (f) show the improvement of the resolution, showing beads which are not resolved in wide-field results.

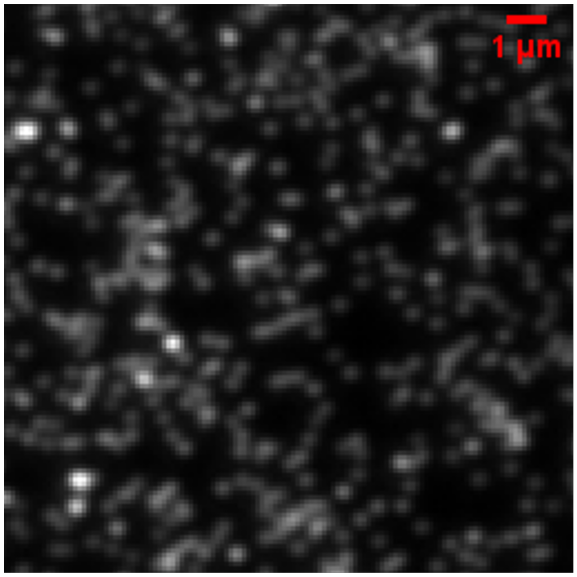
The red-channel results are shwon in Figure 6-19. Similarly, without any correction, the wide-field image (a) and the reconstruction (b) cannot display the entire FOV shown in other cases. When the phase recovery is added, the beads close to each other which cannot be recognised in wide-field (c) are resolved in the reconstructed (d) images. The phase recovery and phase stepping case also demonstrates the resolutiohn enhancement by resolving the beads (f) cannot be clearly distinguished in (e).



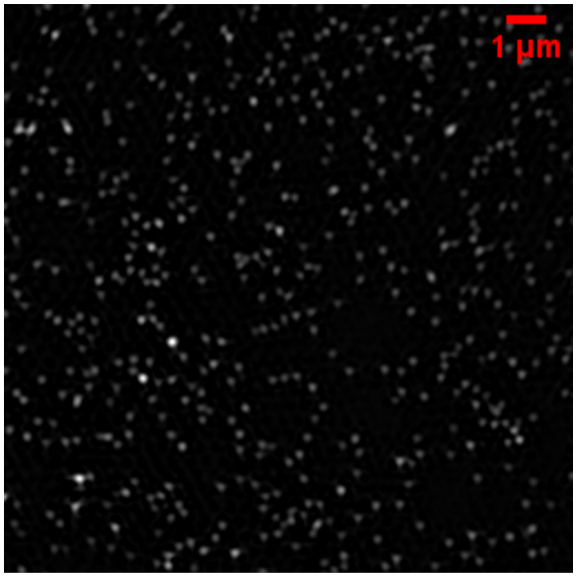
(a)



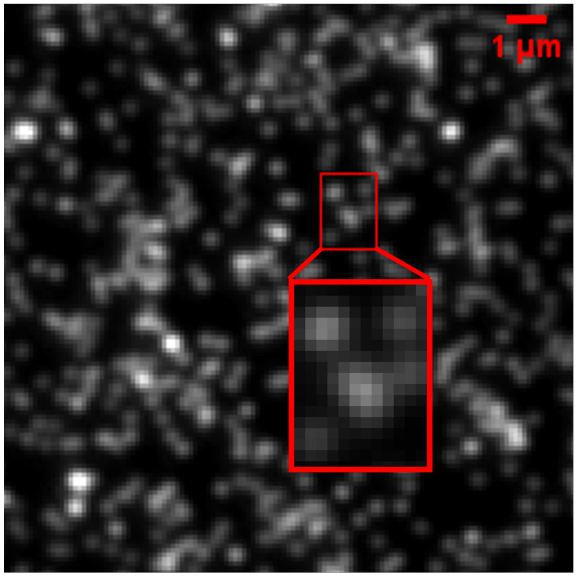
(b)



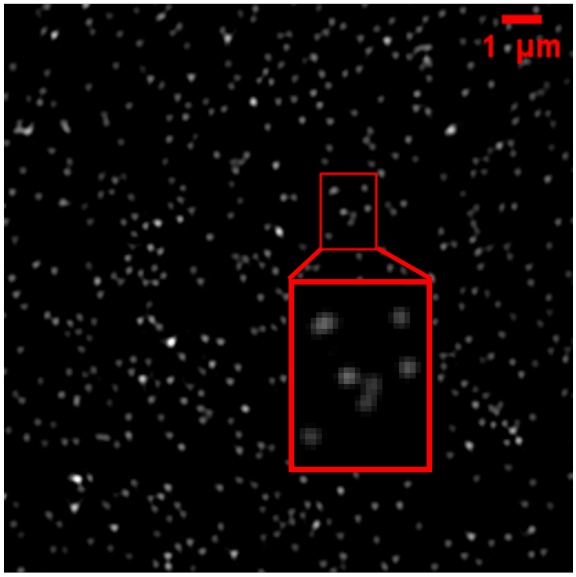
(c)



(d)



(e)



(f)

Figure 6-19. Wide-field images (a), (c), (e) and reconstructed images (b), (d), (f) of 100 nm beads in the red channel (The excitation laser features 561 nm). The first row shows images without any corrections, the second row shows results with phase recovery, and the third row shows results with both phase recovery and phase stepping. All holograms are generated using the DBBI method.

The resolution is determined by measuring the profile of individual beads and fitting the data with a Gaussian function to obtain the full width at half maximum (FWHM) of the system. The reconstructed images in both cases are processed in ImageJ using ThunderSTORM to get the sigma value in the Gaussian fitting. The FWHM values in two channels, as shown in Table 6-2, are calculated as $2.355 \times sigma$. It can be seen that, in both channels, the phase recovery cases have similar FWHM values as the case with phase recovery and phase stepping.

	Phase recovery	Phase recovery & phase stepping
Red channel	166.3 ± 11.3	168.1 ± 11.3
Green channel	138.9 ± 11.8	139.7 ± 11.8

Table 6-2. FWHM of the SIM system. The standard deviation values are taken as the uncertainty. Unit: nm.

The positions of the beams under the two conditions in two channels are found by analysing reconstructed bead-images using ThunderSTORM. The residual artefacts can be analysed by subtracting the fitted gaussians from the reconstructed image data. In this case a uniform gaussian width is used in the subtraction based on the mean of the gaussian width of the fitted gaussians. In order to ignore features that are not successfully fitted by ThunderSTORM only the 9x9 pixel regions of interest around the successful fits are included. The beads that are not detected by ThunderSTORM are hence filtered out. The criteria for success in a fit is that the fitted sigma should be less than (1.4 pixels - which translates to 1.4 * 54 = 75 nm). The residual RMS is calculated by dividing the RMS of the residual image by the RMS of the reconstructed image. As shown in Figure 6-20, in the red channel, the phase recovery case (a) and that with phase recovery and phase stepping (b) have the residual RMS as 37%. In the green channel, the phase-recovery result (c) and the one with additional phase stepping (d) show 39% and 37% RMS, respectively. Overall, the DBBI method with phase correction performs almost the same as the case with additional phase stepping. Nevertheless, introducing phase stepping

can reduce photo-bleaching in certain areas of the biological sample. The results are inconclusive which may be because there are other issues that are producing larger artefacts, such as aberrations, which in this case could (and probably does) include defocus.

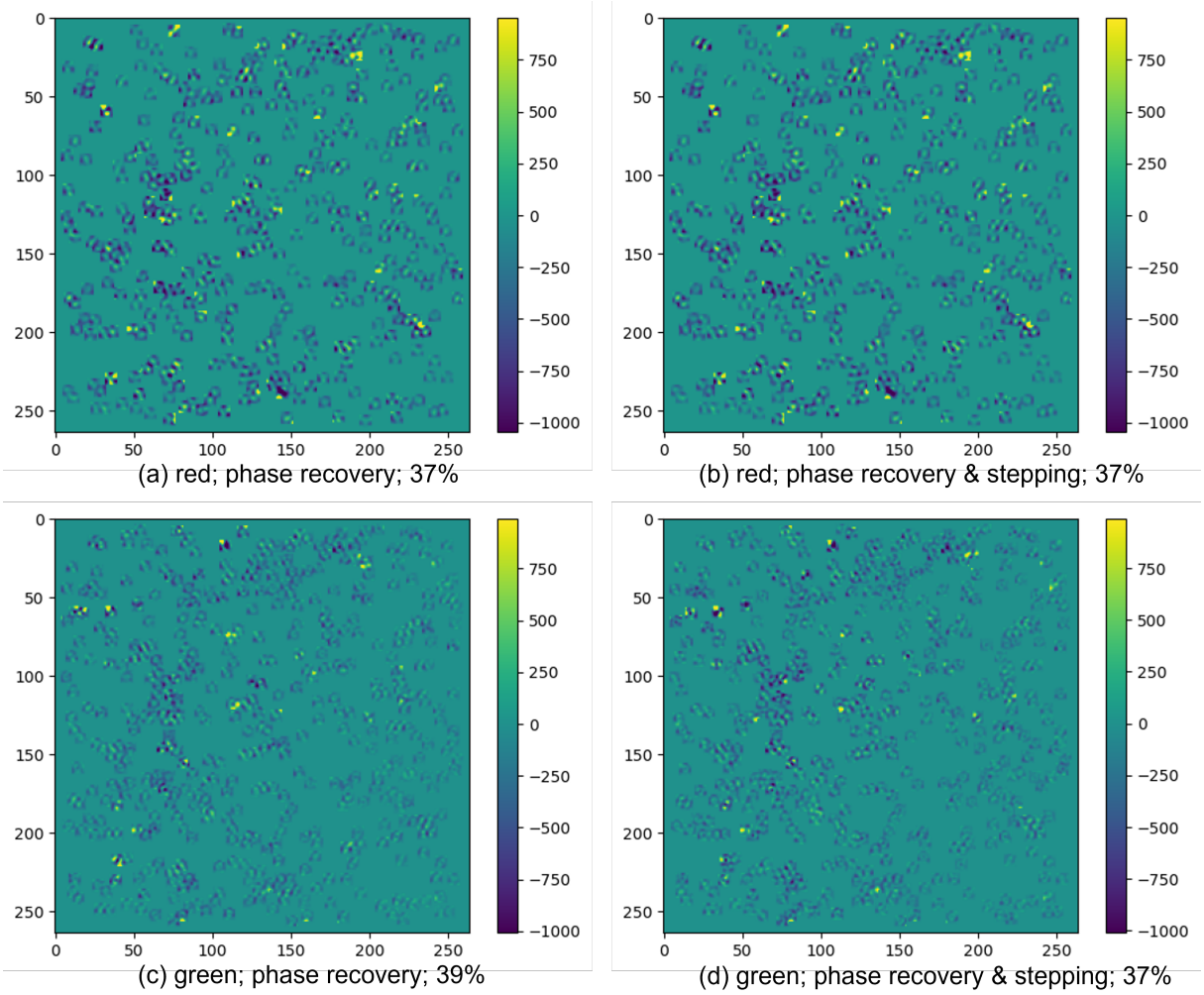


Figure 6-20. Residual results in two channels. The first, second, and third elements in the sub-caption are the imaging channel, the phase situation, and the residual RMS, respectively.

6.5. Chapter summary

In the SLM HexSIM system used by us, both the SLM and the beam splitter placed between the camera for imaging raw SIM images and the objective lens can be accurately and remotely controlled by a piezo controller. A camera is exploited to remotely observe the alignment between diffracted orders and the spatial mask. This makes aligning the SIM system easier and safer to operate from outside the laser safety enclosure. Based on the ScopeFoundry platform in Python, a set of hardware controls and measurement protocols has been built into a project called HexSIMInstrument, that realises all the required equipment. The phase

recovery method introduced in Chapter 4 is implemented on the HexSIM system, where the distorted diffracted orders are measured and corrected in-situ to a diffraction limited state such that they completely pass through pinholes. The SIM imaging is implemented with holograms loaded on the SLM generated in four scenarios (1. The DBBI method is used. 2. The phase stepping is added in the DBBI. 3. The DBBI with phase correction. 4. The DBBI method with phase correction and phase stepping.). Using napari-SIM-processor all sets of holograms show correct calibration from raw images acquired with the microscope. The wide-field images of two cases without phase correction present a limited FOV, while those with phase recovery feature a complete FOV and show more information retrieved from the sample. The raw images are then reconstructed. Compared to the super-resolved images with the DBBI excluding phase correction or stepping, the reconstructed images again show a more complete filed of view and a have clearly different set of reconstructed spatial frequency components. The spectrum of the case with phase correction and phase stepping shows the lowest phase error RMS compared to the cases with only phase correction or no correction at all. Nonetheless, these two results can be compared qualitatively in terms of the uniformity of the brightness within structures imaged within the sample. In both the red and green channels, there are more brighter pixels in the reconstructed images with the DBBI including phase correction than that in the DBBI with phase correction and phase stepping. Fluorescent beads are imaged with different conditions in the two channels. The phase recovery cases have the largest FOVs among results and show clear beads. Although there is no obvious difference between the phase-correction result and that with both phase recovery and phase stepping in terms of the FWHM values and residual results. To summarize, by using the DBBI method with phase correction and phase stepping, the reconstructed SIM can feature a more complete FOV, with less phototoxicity in the sample and fewer artefacts in the final reconstructed image.

Chapter 7. Conclusion and future work

The purpose of this chapter is to summarise the preceding chapters in this thesis and outline directions for future work.

7.1. Conclusion

The focus of this thesis has been on the development and optimisation of a structured illumination microscopy system which uses hexagonal patterns (HexSIM) generated using a ferroelectric liquid crystal spatial light modulator (FLCSLM). A brief introduction of SIM microscopy is in Chapter 2, presenting wide-field microscopy, confocal microscopy and SIM microscopy, revealing the frequency mixing and unmixing for surpassing the diffraction resolution limit and proposing optimisations of the previous SLM HexSIM system.

The initial step of optimising the HexSIM system is to study the performance of the set-up, which is illustrated in Chapter 3. The theory of SIM imaging in generic situations is introduced and is then exploited to simulate practical cases by developing a Python-based software tool: napari-generic-SIMulator. To simulate the imaging system, the plugin can calculate the PSF and OTF using scalar and vectorial theory, respectively. In a vectorial case, since the fluorescent molecules in the sample in a SIM system can be regarded as dipoles, the simulations are implemented with rigid and fixed fluorophores. The PSF and OTF in the scalar situation are similar to the vectorial ones when the polarisation of fluorophores is flexible. While there is a hollow ring in the OTF when the polarisation of fluorescence dipoles is fixed in the same direction of the detection. Then, the structured illumination is calculated based on the theory of Jones vectors. Calculations of conventional 2-D, hexagonal, right-angle hexagonal and conventional 3-D illumination are realised according to the number and position of beams located at the back focal plane of the objective in the SIM system, taking into account the beam polarisations including horizontal, vertical, azimuthal, radial and circular states. In the simulation, the illumination is applied by adding phase titles to each pixel of the sample, and the spatial movement of the specimen is considered. Therefore, raw SIM images can be generated and exploited as the ground truth to test SIM processors and to study different SIM systems in various situations. When running napari-generic-SIMulator, calculations can be accelerated both on CPU and GPU. The stack of raw images can be processed in napari-sim-processor to obtain super-resolved images.

In Chapter 4, the practical problem caused by the non-flat SLM is demonstrated, and then a phase-recovery method is developed to measure and correct the phase aberrations. The method combines the iterative Gerchberg-Saxton algorithm and PSF measurements to overcome the phase singularity and the large phase aberration. The retrieved phase is smoothed using Chebyshev polynomials and is used as feedback to iteratively correct the phase aberration. Sampled Chebyshev expressions are explained, calculating polynomials by interpolating values at some nodal points to each pixel, while weighted polynomials are determined with calculating the orthogonal values. These two approaches both fit the phase and give a rise to similar smooth results. The phase-correction method can be also applied to a round aperture if Zernike polynomials are employed. In both simulations and practical implementations, the phase-recovery method can correct the extended PSF diffracted from the SLM to a diffraction-limited state. To accelerate the iterative phase-recovery algorithm, the holograms are written in the SLM's flash on the fly and the PSF capture is hardware triggered. With the simple optical layout, the distorted PSF is corrected to a diffraction-limited spot after four iterations in less than 20 seconds. The last measured wavefront RMS is 0.063 radians, which is evidently smaller than the diffraction limited value of 0.469 radians.

In Chapter 5, three methods to design hexagonal holograms are presented and compared: the polka-dots approximation, the method of direct binarisation of beam interference (DBBI) and the optimisation by direct binary search. The polka-dots approach is exploited to draw holograms according to the geometric shapes of the dots, which is the conventional way for generating hexagonal patterns. The DBBI method is generic to generate structured patterns based on the interference between beams, which can include the phase correction and control the phases between beams and can be extended to generate patterns of different numbers of beams. The optimisation of binary search can be adapted to recover the phase aberration caused by the SLM but is too slow to be an appropriate approach. In situations with and without the phase aberration, the DBBI method reaches the highest diffraction efficiency than the polka-dots approximation with the ratio of dots and non-dots area as 4:6 and the optimisation by binary search, where the 5:5 polka dots show the lowest diffraction intensity. The phase aberration causes the incomplete field of the diffraction, and this issue can be solved using the DBBI method. Additionally, DBBI has the highest tolerance of the alignment between three beams and the mask than the other two approaches of producing holograms. In fact, the method of 4:6 polka dots performs very similar to the optimisation of binary search, which is thus not discussed because its slow speed. In terms of the amplitude and phase of carriers used to determine SIM reconstruction factors, applying stepped phases in DBBI can decrease the fluctuation of carriers' values to provide smaller standard deviations. With the phase

aberration retrieved in our system, the pinhole size cannot be minimised when 4:6 dots approximation is applied, whereas the size can be halved if DBBI with the phase correction is exploited, although compromising the tolerance of the misalignment. The pinhole can be enlarged twice to make the aberrated beams pass when the 4:6 dots method is employed, however, more higher orders can go through the pinholes as well, making the parameters of the carriers less accurate. Therefore, the DBBI with the phase recovery and the phase stepping is a decent method to generate holograms for SIM illumination.

In Chapter 6, the implementation of our HexSIM is presented. Some components are introduced into the optical layout to realise the alignment of the system remotely, easily and conveniently, including the optics for the phase recovery implementation, the piezo stage and actuator and the extra camera to image the mask. The previous ScopeFoundry based programme: HexSIMInstrument is improved to integrate all the hardware and measurements. The applications of phase correction and phase stepping in the DBBI method are presented. The camera used to image the sample and the SLM exploited to display holograms are synced by the hardware triggering. The phase aberration resulted from the non-flat SLM is recovered by applying the in-situ phase correction method, the distorted diffracted orders can hence entirely pass through the pinholes. The resultant wavefront RMS is 0.113 radians, which is significantly smaller than the diffraction limited value of 0.469 radians. Measurements are carried out on a fluorescent sample with using the DBBI method with phase recovery and phase stepping. The raw images obtained with the case using the phase recovery show an obviously larger FOV than results without the phase correction. Furthermore, the reconstructed image in the situation with the DBBI only including phase correction show more brighter pixels than cases with both the phase recovery and phase stepping. The with-correction spectrum is of clearer frequency component than the no-correction spectrum. The phase-stepping result has less blurs around the carrier frequencies than that without phase stepping. Apart from the improvement in determining the carriers, applying the BBI method with phase stepping can practically reduce the possibility of photo-bleaching at certain areas in the sample. The DBBI method with phase correction and phase stepping is a superior approach for SIM imaging.

7.2. Future work

The work discussed in the thesis can be extended to further develop and optimise SIM systems. In terms of more automatic control of the sample location, the x-y and z stage can be integrated into the HexSIMInstrument and synchronised to hologram display on the SLM to obtain stacks of raw SIM images with 3-D scanning.

The control of the polarisations of beams diffracted from the SLM can be added by introducing polarisers at the mask plane as was originally the case with the system before this work was started. Further simulations using napari-generic-SIMulator can be used to analyse and optimise this approach. It is noteworthy that the usual approach of using azimuthal polarisation for HexSIM inverts the structured illumination contrast relative to that of linear polarisation at low NA, with a subsequent halving of the amplitude of the carrier terms. Clegg presented an efficient method of introducing full polarisation control with a ferro-electric liquid crystal SLM in a double pass arrangement [64], though with a double pass arrangement where two holograms are displayed on the SLM. It may be beneficial to investigate this approach for SIM, or even to use two separate SLMs in sequence to give full polarisation control to the diffracted beams.

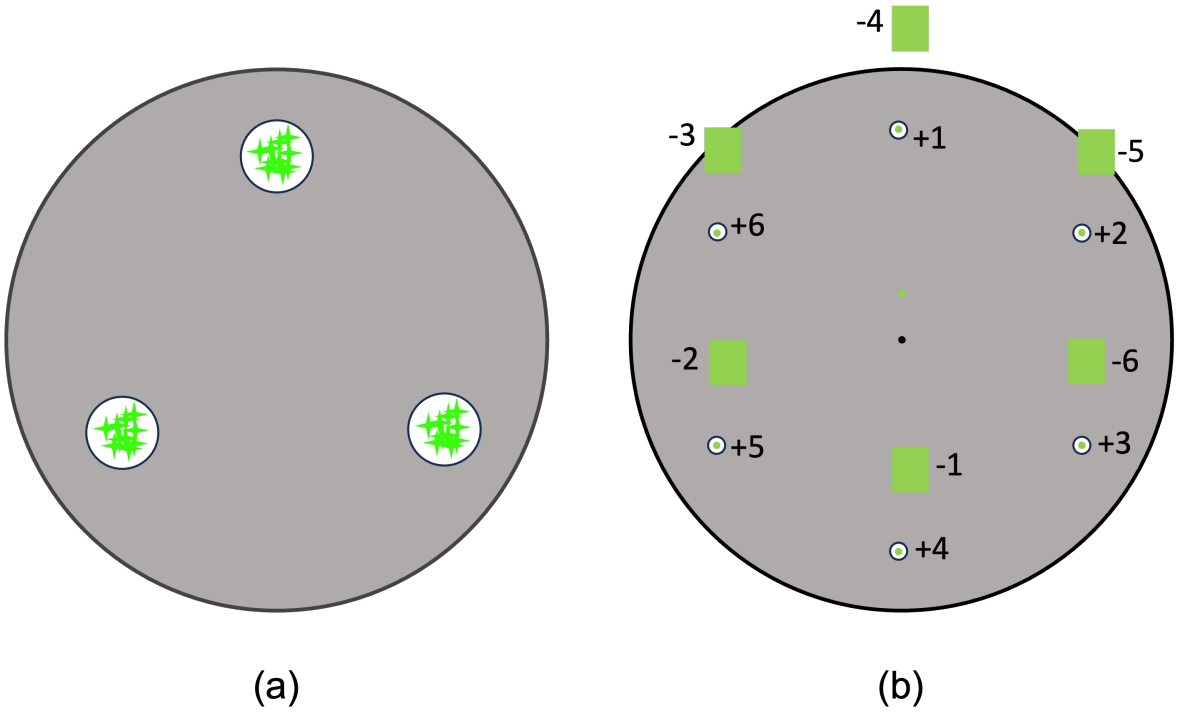


Figure 7-1. (a) Schematic of diffracted orders with hologram beam shifting. A single beam (denoted by a green star) is diffracted into each pinhole by each hologram in a sequence that dithers the spot positions. (b) Schematic when applying the DBBI with correction in a conventional SIM geometry with sinusoidal pattern illumination. The centre of the diffracted orders (the optic axis) is denoted in green and that of the mask in black.

In his seminal 2008 paper, Gustaffson also showed that by using a multimode fibre to decrease spatial coherence, the contrast in the structed illumination can be reduced in its axial extent, with additional reduced sensitivity of the generated patterns to undesired interference resulting from dust or stray reflections [27], another potential source of artefacts. Based on holograms

with stepped phases, the spatial coherence of illuminating beams can be similarly reduced. While phase stepping in the current HexSIM system, a small shift could be introduced at each hologram. Consequently, as shown in Figure 7-1 (a), each desired diffracted order passing through the pinhole would consist of ten diffraction limited beams produced by the holograms displayed sequentially on the SLM (when the phase step is set as ten). The separation of the beams is always the same, so in-focus the interference is also the same. However, out of focus the pattern will move corresponding to the shift in the input beams. This approach is also likely to have a beneficial effect on artefacts produced by the interference with aliased higher diffracted orders, in a similar way that the phase-stepping approach presented in chapter 5 does.

The DBBI method with phase correction and phase stepping can be applied in a conventional SIM system using sinusoidal illumination patterns. However, as demonstrated in Chapter 6, when applying a flatness correction to the hologram, the conjugate of a corrected diffracted order has twice the phase aberration. To guarantee all diffracted beams passing through pinholes are diffraction limited, each hologram needs to diffract two beams (four beams if conjugates are considered), and there should be a shift in the location of the diffracted beams with respect to the optic-axis to avoid the beams overlapping at the mask. As presented in Figure 7-1 (b), each diffracted beam passing through a pinhole (green spot in the circle) has a conjugate (showed as a large rectangular since it has the twice aberration) that is blocked by the mask. All wanted diffracted orders are denoted with positive numbers, while the unwanted conjugates are numbered with a negative sign. Obviously, the distance between ± 1 order to the optic axis (the green centre) is smaller than that between ± 4 order and the optic axis. Similar, the distance of ± 2 or ± 6 is smaller than that of ± 5 and ± 3, respectively. The different positions of the centre of the diffraction (the green centre at the optic axis) and that of the mask (the black centre) show the shift of the alignment between the SLM optic axis and the mask. Using this approach we can now also provide diffracted light to the mask centre position and thereby implement a classic 3-beam, 3-D SIM geometry with correction applied at the new DC order.

Using this approach, the whole system can be extended to support 3-D SIM. In fact, regarding three dimensions, napari-generic-SIMulator is capable to simulate raw images for 3-D SIM geometries, although napari-SIM-processor requires to be refined to properly calibrate 3-D data.

Chapter 3 in this thesis describes the simulation of SIM data including an idealised illumination from a set of polarised beams. This simulation is then presented as a freely available plugin to the napari system, and indeed it has been very helpful in the development of our own processing plugin. However, Chapter 5 then goes on to demonstrate the sort of artefacts that are present in the illumination when a pixellated SLM is used to generate the required beams by diffraction. It would be useful to be able to integrate this same approach into the napari-generic-SIMulator plugin. This is not straightforward, however, as the analysis in Chapter 5 effectively calculates the illumination intensity in the intermediate image plane of the microscope where the numerical aperture is low, and a purely scalar approach is sufficient. Here too, fast Fourier transforms are used but these only calculate the intensity pattern at a discrete set of data-points on a regular grid. To be useful in the full SIMulator we would need to be able to use a full vectorial calculation to work out the polarised intensities in the sample. The sample in the SIMulator is also a point cloud where the point positions are by design located arbitrarily in 3-D space, and so a regular grid calculation of the illumination intensity in a 2-D plane is not so useful. Currently the SIMulator performs a point-by-point calculation of the illumination intensity from the input polarised beams, which is very fast as there are only a small number of beams in the system. The challenge of including the SLM simulation into the code will therefore be to find a way of calculating the sample illumination at arbitrary points in the sample on a point-by-point basis as efficiently as possible, when the input consists of a 2-D array of SLM pixels.

Overall, the software frameworks and hardware developed in this thesis provide excellent opportunities for on-going research into SIM at Imperial, just as the previous work on SIM at Imperial has provided for the work presented here in this thesis.

References

1. D. J. Webb and C. M. Brown, "Epi-Fluorescence Microscopy," in (2012), pp. 29–59.
2. G. G. Stokes, "XXX. On the change of refrangibility of light," Philos. Trans. R. Soc. London **142**, 463–562 (1852).
3. "BeitrÃge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung," Arch. für mikroskopische Anat. **9**(1), 418–440 (1873).
4. J. B. Pawley, ed., *Handbook Of Biological Confocal Microscopy* (Springer US, 2006).
5. M. Minsky, "Microscopy Apparatus," U.S. patent 3013467 (November 7, 1957).
6. K. R. Spring, T. J. Fellers, and M. W. Davidson, "Confocal Microscopy - Signal-to-Noise considerations," https://www.olympus-lifescience.com/en/microscope-resource/primer/techniques/confocal/signaltonoise/.
7. A. Egner, V. Andresen, and S. W. Hell, "Comparison of the axial resolution of practical Nipkow-disk confocal fluorescence microscopy with that of multifocal multiphoton microscopy: theory and experiment," J. Microsc. **206**(1), 24–32 (2002).
8. A. D. Elliott, "Confocal Microscopy: Principles and Modern Practices," Curr. Protoc. Cytom. **92**(1), (2020).
9. M. Petráň, M. Hadravský, M. D. Egger, and R. Galambos, "Tandem-Scanning Reflected-Light Microscope*," J. Opt. Soc. Am. **58**(5), 661 (1968).
10. R. Rottenfusser, E. E. Wilson, and M. W. Davidson, "The Point Spread Function," https://zeiss-campus.magnet.fsu.edu/articles/basics/psf.html.
11. M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, "Sevenfold improvement of axial resolution in 3D wide-field microscopy using two objective lenses," in T. Wilson and C. J. Cogswell, eds. (1995), pp. 147–156.
12. N. Streibl, "Three-dimensional imaging by a microscope," J. Opt. Soc. Am. A **2**(2), 121 (1985).
13. D. A. Agard, Y. Hiraoka, P. Shaw, and J. W. Sedat, *Fluorescence Microscopy in Three Dimensions. Methods in Cell Biology* (1989), **30**(27).
14. M. G. L. Gustafsson, "Nonlinear structured-illumination microscopy: Wide-field

fluorescence imaging with theoretically unlimited resolution," Proc. Natl. Acad. Sci. U. S. A. **102**(37), 13081–13086 (2005).

15. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," Nat. Methods **3**(10), 793–795 (2006).

16. M. Heilemann, S. Van De Linde, M. Schüttpelz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, and M. Sauer, "Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes," Angew. Chemie - Int. Ed. **47**(33), 6172–6176 (2008).

17. A. Lampe, V. Haucke, S. J. Sigrist, M. Heilemann, and J. Schmoranzer, "Multi-colour direct STORM with red emitting carbocyanines," Biol. Cell **104**(4), 229–237 (2012).

18. S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," Opt. Lett. **19**(11), 780 (1994).

19. T. A. Klar and S. W. Hell, "Subdiffraction resolution in far-field fluorescence microscopy," Opt. Lett. **24**(14), 954 (1999).

20. T. J. Gould, L. K. Schroeder, P. A. Pellett, and J. Bewersdorf, "STED Microscopy," Fluoresc. Microsc. From Princ. to Biol. Appl. Second Ed. 321–338 (2017).

21. M. A. A. Neil, R. Juškaitis, and T. Wilson, *Method of Obtaining Optical Sectioning by Using Structured Light in a Conventional Microscope* (1997), **22**(24).

22. R. Heintzmann and T. Huser, "Super-Resolution Structured Illumination Microscopy," Chem. Rev. **117**(23), 13890–13908 (2017).

23. M. A. A. Neil, R. Juškaitis, and T. Wilson, "Real time 3D fluorescence microscopy by two beam interference illumination," Opt. Commun. **153**(1–3), 1–4 (1998).

24. D. Karadaglić and T. Wilson, "Image formation in structured illumination wide-field fluorescence microscopy," Micron **39**(7), 808–818 (2008).

25. B. Bailey, V. Krishnamurthi, D. L. Farkas, D. L. Taylor, and F. Lanni, "Three-dimensional imaging of biological specimens with standing wave fluorescence microscopy," SPIE **2184**, 208–213 (1994).

26. R. Heintzmann and C. G. Cremer, "Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating," Opt. Biopsies Microsc. Tech. III **3568**(January 1999), 185–196 (1999).

27. M. G. L. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," J. Microsc. **198**(2), 82–87 (2000).

28. M. G. L. Gustafsson, L. Shao, P. M. Carlton, C. J. R. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat, "Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination," Biophys. J. **94**(12), 4957–4970 (2008).

29. M. G. L. Gustafsson, "Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution.," Proc. Natl. Acad. Sci. U. S. A. **102**(37), 13081–6 (2005).

30. S. A. Shroff, J. R. Fienup, and D. R. Williams, "OTF compensation in structured illumination superresolution images," Unconv. Imaging IV **7094**, 709402 (2008).

31. M. Piana and M. Bertero, "Regularized deconvolution of multiple images of the same object," J. Opt. Soc. Am. A **13**(7), 1516 (1996).

32. A. Lal, C. Shan, and P. Xi, "Structured Illumination Microscopy Image Reconstruction Algorithm," IEEE J. Sel. Top. Quantum Electron. **22**(4), 50–63 (2016).

33. M. Schropp and R. Uhl, "Two-dimensional structured illumination microscopy," J. Microsc. **256**(1), 23–36 (2014).

34. E. A. Ingerman, R. A. London, R. Heintzmann, and M. G. L. Gustafsson, "Signal, noise and resolution in linear and nonlinear structured-illumination microscopy," J. Microsc. **273**(1), 3–25 (2019).

35. L. Amit, S. Chunyan, and X. Peng, "Structured Illumination Microscopy image reconstruction algorithm," IEEE J. Sel. Top. QUANTUM Electron. **22**(4), (2016).

36. J. Demmerle, C. Innocent, A. J. North, G. Ball, M. Müller, E. Miron, A. Matsuda, I. M. Dobbie, Y. Markaki, and L. Schermelleh, "Strategic and practical guidelines for successful structured illumination microscopy," Nat. Protoc. **12**(5), 988–1010 (2017).

37. M. Schropp, C. Seebacher, and R. Uhl, "XL-SIM: Extending superresolution into deeper layers," Photonics **4**(2), (2017).

38. H. Gong, W. Guo, and M. A. A. Neil, "GPU-accelerated real-time reconstruction in Python of three-dimensional datasets from structured illumination microscopy with hexagonal patterns," Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **379**(2199), 20200162 (2021).

39. X. Chen, S. Zhong, Y. Hou, D. Kim, and P. Xi, "Superresolution structured illumination microscopy reconstruction algorithms : a review," 1–34 (2023).

40. E. H. Rego, L. Shao, J. J. Macklin, L. Winoto, G. A. Johansson, N. Kamps-Hughes, M. W. Davidson, and M. G. L. Gustafsson, "Nonlinear structured-illumination microscopy

with a photoswitchable protein reveals cellular structures at 50-nm resolution," Proc. Natl. Acad. Sci. U. S. A. **109**(3), (2012).

41. H. W. Lu-Walther, W. Hou, M. Kielhorn, Y. Arai, T. Nagai, M. M. Kessels, B. Qualmann, and R. Heintzmann, "Nonlinear structured illumination using a fluorescent protein activating at the readout wavelength," PLoS One **11**(10), 1–14 (2016).

42. B. Liu, C. Chen, X. Di, J. Liao, S. Wen, Q. P. Su, X. Shan, Z. Q. Xu, L. A. Ju, C. Mi, F. Wang, and D. Jin, "Upconversion Nonlinear Structured Illumination Microscopy," Nano Lett. **20**(7), 4775–4781 (2020).

43. K. Wicker and R. Heintzmann, "Resolving a misconception about structured illumination," **8**(May), (2014).

44. D. Li, L. Shao, B. C. Chen, X. Zhang, M. Zhang, B. Moses, D. E. Milkie, J. R. Beach, J. A. Hammer, M. Pasham, T. Kirchhausen, M. A. Baird, M. W. Davidson, P. Xu, and E. Betzig, "Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics," Science (80-.). **349**(6251), (2015).

45. M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, "Multicolor super-resolution imaging with photo-switchable fluorescent probes," Science (80-.). **317**(5845), 1749–1753 (2007).

46. S. H. Shim, C. Xia, G. Zhong, H. P. Babcock, J. C. Vaughan, B. Huang, X. Wang, C. Xu, G. Q. Bi, and X. Zhuang, "Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes," Proc. Natl. Acad. Sci. U. S. A. **109**(35), 13978–13983 (2012).

47. S. J. Holden, S. Uphoff, and A. N. Kapanidis, "DAOSTORM: An algorithm for high-density super-resolution microscopy," Nat. Methods **8**(4), 279–280 (2011).

48. S. A. Jones, S. H. Shim, J. He, and X. Zhuang, "Fast, three-dimensional super-resolution imaging of live cells," Nat. Methods **8**(6), 499–505 (2011).

49. D. B. Murphy and M. W. Davidson, *Fundamentals of Light Microscopy and Electronic Imaging* (John Wiley & Sons, Incorporated, 2012).

50. M. Lelek, M. T. Gyparaki, G. Beliu, F. Schueder, J. Griffié, S. Manley, R. Jungmann, M. Sauer, M. Lakadamyali, and C. Zimmer, "Single-molecule localization microscopy," Nat. Rev. Methods Prim. **1**(1), 39 (2021).

51. J. Lippincott-Schwartz, "Profile of Eric Betzig, Stefan Hell, and W. E. Moerner, 2014 Nobel Laureates in Chemistry," Proc. Natl. Acad. Sci. **112**(9), 2630–2632 (2015).

52. T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell, "Fluorescence microscopy

with diffraction resolution barrier broken by stimulated emission.," Proc. Natl. Acad. Sci. U. S. A. **97**(15), 8206–10 (2000).

53. M. Hofmann, C. Eggeling, S. Jakobs, and S. W. Hell, "Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins," Proc. Natl. Acad. Sci. U. S. A. **102**(49), 17565–17569 (2005).

54. T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell, "Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission," Proc. Natl. Acad. Sci. **97**(15), 8206–8210 (2000).

55. B. R. Rankin, G. Moneron, C. A. Wurm, J. C. Nelson, A. Walter, D. Schwarzer, J. Schroeder, D. A. Colón-Ramos, and S. W. Hell, "Nanoscopy in a Living Multicellular Organism Expressing GFP," Biophys. J. **100**(12), L63–L65 (2011).

56. T. napari Team, "napari: a fast, interactive viewer for multi-dimensional images in Python — napari," https://napari.org/stable/.

57. M. Liang, "napari generic SIMulator," https://www.napari-hub.org/plugins/napari-generic-SIMulator.

58. A. Bassi and M. Neil, "napari SIM processor acceleration," (n.d.).

59. K. O'Holleran and M. Shaw, "Polarization effects on contrast in structured illumination microscopy," Opt. Lett. **37**(22), 4603 (2012).

60. M. Born and E. Wolf, *Principles of Optics .Pdf* (Pergamon Press, 1980).

61. V. Ignatowsky, "Diffraction by a lens of arbitrary aperture," Trans. Opt. Inst. Petrograd. **1**(4), 1 (1919).

62. V. Ignatowsky, "The relationship between geometrical and wave optics and diffraction of homocentrical beams," Trans. Opt. Inst. **1**, 1 (1920).

63. T. Computing, "Electromagnetic diffraction in optical systems, II. Structure of the image field in an aplanatic system," Proc. R. Soc. London. Ser. A. Math. Phys. Sci. **253**(1274), 358–379 (1959).

64. J. H. Clegg, "Polarisation microscopy of single emitters," (2015).

65. B. R. Boruah and M. A. A. Neil, "Focal field computation of an arbitrarily polarized beam using fast Fourier transforms," Opt. Commun. **282**(24), 4660–4667 (2009).

66. M. J. Booth, M. A. A. Neil, R. Juškaitis, and T. Wilson, "Adaptive aberration correction in a confocal microscope," Proc. Natl. Acad. Sci. U. S. A. **99**(9), 5788–5792 (2002).

67. D. Débarre, E. J. Botcherby, T. Watanabe, S. Srinivas, M. J. Booth, and T. Wilson,

"Image-based adaptive optics for two-photon microscopy," **34**(16), 2495–2497 (2009).

68. M. Rueckel, J. A. Mack-Bucher, and W. Denk, "Adaptive wavefront correction in two-photon microscopy using coherence-gated wavefront sensing," Proc. Natl. Acad. Sci. U. S. A. **103**(46), 17137–17142 (2006).

69. P. Marsh, D. Burns, and J. Girkin, "Practical implementation of adaptive optics in multiphoton microscopy," Opt. Express **11**(10), 1123 (2003).

70. N. Olivier, D. Débarre, and E. Beaurepaire, "Dynamic aberration correction for multiharmonic microscopy," Opt. Lett. **34**(20), 3145–3147 (2009).

71. D. Débarre, T. Vieille, and E. Beaurepaire, "Simple characterisation of a deformable mirror inside a high numerical aperture microscope using phase diversity," J. Microsc. **244**(2), 136–143 (2011).

72. A. Facomprez, E. Beaurepaire, and D. Débarre, "Accuracy of correction in modal sensorless adaptive optics," Opt. Express **20**(3), 2598 (2012).

73. M. J. Booth, "Adaptive optical microscopy: The ongoing quest for a perfect image," Light Sci. Appl. **3**(January), 1–7 (2014).

74. D. Burke, B. Patton, F. Huang, J. Bewersdorf, and M. J. Booth, "Adaptive optics correction of specimen-induced aberrations in single-molecule switching microscopy," Optica **2**(2), 177 (2015).

75. B. M. Hanser, M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, "Phase retrieval for high-numerical-aperture optical systems," Opt. Lett. **28**(10), 801 (2003).

76. D. Turaga and T. E. Holy, "Image-based calibration of a deformable mirror in wide-field microscopy," Appl. Opt. **49**(11), 2030–2040 (2010).

77. D. Debarre, M. J. Booth, and T. Wilson, "Image based adaptive optics through optimisation of low spatial frequencies," Opt. Express **15**(13), 8176 (2007).

78. W. Lukosz, "Der Einfluß der Aberrationen auf die Optische Übertragungsfunktion bei Kleinen Orts-Frequenzen," Opt. Acta Int. J. Opt. **10**(1), 1–19 (1963).

79. M. G. Löfdahl, G. B. Scharmer, and W. Wei, "Calibration of a deformable mirror and Strehl ratio measurements by use of phase diversity," Appl. Opt. **39**(1), 94 (2000).

80. M. Booth, T. Wilson, H. Sun, T. Ota, and S. Kawata, "Methods for the characterization of deformable membrane mirrors," Appl. Opt. **44**(24), 5131–5139 (2005).

81. D. Débarre, E. J. Botcherby, M. J. Booth, and T. Wilson, "Adaptive optics for structured illumination microscopy," Opt. Express **16**(13), 9290 (2008).

82. M. Schwertner, M. J. Booth, M. A. A. Neil, and T. Wilson, "Measurement of specimen-induced aberrations of biological samples using phase stepping interferometry," J. Microsc. **213**(1), 11–19 (2004).

83. B. M. Hanser, M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, "Phase-retrieved pupil functions in wide-field fluorescence microscopy," J. Microsc. **216**(1), 32–48 (2004).

84. C. Ji, Y. Zhu, E. He, Q. Liu, D. Zhou, S. Xie, H. Wu, J. Zhang, K. Du, Y. Chen, W. Liu, and C. Kuang, "Full field-of-view hexagonal lattice structured illumination microscopy based on the phase shift of electro–optic modulators," Opt. Express **32**(2), 1635 (2024).

85. R. Heintzmann, T. M. Jovin, and C. Cremer, "Saturated patterned excitation microscopy—a concept for optical resolution improvement," J. Opt. Soc. Am. A **19**(8), 1599 (2002).

86. R. Heintzmann, "Saturated patterned excitation microscopy with two-dimensional excitation patterns," Micron **34**(6–7), 283–291 (2003).

87. L. Hirvonen, O. Mandula, K. Wicker, and R. Heintzmann, "Structured illumination microscopy using photoswitchable fluorescent proteins," Three-Dimensional Multidimens. Microsc. Image Acquis. Process. XV **6861**, 68610L (2008).

88. P. Kner, B. B. Chhun, E. R. Griffis, L. Winoto, and M. G. L. Gustafsson, "Super-resolution video microscopy of live cells by structured illumination," Nat. Methods **6**(5), 339–342 (2009).

89. K. Temma, R. Oketani, T. Kubo, K. Bando, S. Maeda, K. Sugiura, T. Matsuda, R. Heintzmann, T. Kaminishi, K. Fukuda, M. Hamasaki, T. Nagai, and K. Fujita, "Selective-plane-activation structured illumination microscopy," Nat. Methods (2024).

90. M. T. M. Hannebelle, E. Raeth, S. M. Leitao, T. Lukeš, J. Pospíšil, C. Toniolo, O. F. Venzin, A. Chrisnandy, P. P. Swain, N. Ronceray, M. P. Lütolf, A. C. Oates, G. M. Hagen, T. Lasser, A. Radenovic, J. D. McKinney, and G. E. Fantner, "Open-source microscope add-on for structured illumination microscopy," Nat. Commun. **15**(1), 1–8 (2024).

91. M. L. Senftleben, A. Bajor, E. Hirata, S. Abrahamsson, and H. Brismar, "Fast volumetric multifocus structured illumination microscopy of subcellular dynamics in living cells," Biomed. Opt. Express **15**(4), 2281 (2024).

92. P. G. de Gennes and J. Prost, *The Physics of Liquid Crystals*, Second (Oxford Science Publications, 1995).

93. G. Andersson, I. Dahl, P. Keller, S. T. Lagerwall, K. Skarp, and B. Stebler,

"Submicrosecond electro - optic switching in the liquid - crystal smectic A phase: The soft-mode ferroelectric effect," **642**, 640–642 (1987).

94. R. B. MEYER, L. LIÉBERT, L. STRZELECKI, and P. KELLER, "Ferroelectric liquid crystals," J. Phys. Lett. (36), 69–71 (1975).

95. S. E. Broomfield, M. A. A. Neil, E. G. S. PaigE, and G. G. Yang, "Programmable binary phase-only optical device based on ferroelectric liquid crystal SLM," **28**(1), (1992).

96. M. Liang, "Structured illumination microscope - Thesis supplementary videos," https://figshare.com/articles/media/Structured_Illumination_Microscopy_-_Thesis_supplementary_videos/26125087.

97. M. A. Seldowitz, J. P. Allebach, and D. W. Sweeney, "Synthesis of digital holograms by direct binary search," Appl. Opt. **26**(14), 2788 (1987).

98. M. P. Dames, R. J. Dowling, P. Mckee, and D. Wood, "Efficient optical elements to generate intensity weighted spot arrays : design and fabrication," **30**(19), 2685–2691 (1991).

99. M. Žurauskas, "Prototype Optical System for Aperture Correlation Microscopy Device," Imperial College London (2010).

100. The Qt. Company, "Qt Documentation," https://doc.qt.io/.