THE UNIVERSITY OF HONG KONG

DOCTORAL THESIS

Metasurface for "Vision" improvement

Author:
Huade MAO

Supervisor:
Prof. Kenneth K.Y. WONG
Co-Supervisor:
Prof. Kevin K.M. TSIA

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

in the

Department of Electrical and Electronic Engineering
Faculty of Engineering
The University of Hong Kong

February 20, 2024

Abstract of thesis entitled

Metasurface for "Vision" improvement

Submitted by

Huade MAO

for the degree of Doctor of Philosophy

at The University of Hong Kong

in February, 2024

Metasurface, a miniaturized device with a meticulously designed permittivity distribution which accommodates nano-fabrication technique, has many novel optical functionalities. This thesis mainly presents the design and application of metasurface devices to improve "vision" in three aspects, namely, signal "vision", holographic "vision", and biomedical "vision". The design techniques spread from forward calculation, neural network, to adjoint optimization. These methods are dedicated to finalize the exact metasurface patterning for a desired far-field radiation, which, if properly defined, can be beam profiles in signal "vision", 2D intensity distribution in holographic "vision", or depth of focus distribution in biomedical "vision".

For signal "vision", an enhanced signal transmission is achieved through a metasurface to convert a plane wave into multiple Laguerre-Gaussian (LG) modes, thus expanding the communication channels by one more dimension. A 2D array of Au nano-antenna was fabricated on a silica chip. When the chip was illuminated by a plane wave, multiple orthogonal LG modes will be generated and demultiplexed into different angles. Each nano-antenna has a width of 200 nm and a length of 200-400 nm, which can achieve a complex modulation with 0-1 amplitude range and 0-2π phase range. Error analysis suggests that this metasurface is robust for LG mode transmission over a 400-nm broadband range.

For holographic "vision", the arbitrary holographic display is accomplished through a metasurface predicted by a pre-trained neural network in a lensless projection modality. The design process through neural network is about 400 times faster than theoretical computation. The metasurface in question is capable of complex modulation. Here I developed a convolutional neural network with dual-output, featuring the complex pattern in terms of the amplitude and phase branch. Then the complex pattern, termed as computer generated hologram (CGH) was fabricated into a metasurface and positioned in an experimental setup to project a holographic image. The averaged training loss was 4.43% while the testing loss was 4.57%, suggesting a good generalization.

For biomedical "vision", an augmented biomedical imaging can be realized by a tailored fiber facet, which is inversely designed through adjoint optimization to deliver an

extended depth of focus for blood vessel imaging in optical coherence tomography. Two modalities has been proposed. One is chip-modality: a phase mask has been optimized to deliver an extended focus with a center intensity 40% higher than its Bessel beam counterparts and a lowered side lobe. The other is fiber -modality: a grating structure has been calculated on top of a fiber facet to an extended focus in the forward direction. In the future, I will calculate the grating pattern for a fiber with side view and integrate it to an imaging modality.

All in all, metasurface design techniques have been demonstrated to improve "vision" in signal, holographic, and biomedical aspects. The design techniques spread from forward calculation for a complex map of signal demultiplexing, neural neural network prediction for an accelerated holographic display, and all the way to the inverse design achieved by adjoint optimization to finalize the design in one go to meet the desired requirement.

(An abstract of 498 words)

Metasurface for "Vision" improvement

by

Huade MAO
B.S. *Nanjing University*

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy

at

The University of Hong Kong
February, 2024

Declaration

I, Huade MAO, declare that this thesis titled, “Metasurface for "Vision" improvement”, which is submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy, represents my own work except where due acknowledgement have been made. I further declared that it has not been previously included in a thesis, dissertation, or report submitted to this University or to any other institution for a degree, diploma or other qualifications.

Signed: Huade MAO

Date: 2024-Feb-23

Acknowledgements

I started my Ph.D. at HKU in the autumn of 2019. It was several years ago, but it feels like a blink of time. One minute I was still a new student looking forward to exploring the science, the next I am preparing the thesis and marking the closure of my academia life. Along the Ph.D. journey, I've seen and done a lot, all of which are connected together to create a unique memory. Up to now, I think I owe lots of people or parties a great debt of gratitude upon the completion of this thesis.

Firstly, I would like to thank my supervisor, Prof. Kenneth K.Y. Wong, who gave me an opportunity to start my Ph.D. research in his group. Kenneth has given me lots of flexibility in terms of my research topic. By then, I was interested in metasurface, a subject our group never touched before. But Kenneth was being super supportive. He bought the device I needed and checked my progress on a bi-weekly basis. Without his motivation, I could never go this far in metasurface field.

Then, my parents are definitely the most important people in my life I should express my gratitude to. We are bonded together by the strongest tie in the world - family. There are lots of arguments between me and my parents. For example, my parents were very against the idea of pursuing a Ph.D. education outside Mainland, China. But in the end, they began to respect my decision. I fully appreciate their support throughout my whole Ph.D. life.

I also want to give blessing to this wonderful city - Hong Kong. This city gives me a chance to get my Ph.D. education, which also allows me to travel out of Mainland China for the first time. The 4-year Ph.D. journey is not only intellectual inspiring, but also cross-culture immersive. Hong Kong is vibrant and international. I've been able to network with lots of professionals from all over the world.

Next, I shall give my sincere gratitude to the lab mates I have been working with. Their companion consists of an important part of my Ph.D. life. They are Jiqiang KANG, Yitian TONG, Tian QIAO, Chandra JINATA, Pingping FENG, Jiawei SHI, Sheng WANG, Hongsen HE, Yi ZHOU, Ryan Ka Yan CHAN, Xiaomeng CUI, Najia SHARMIN, Meng ZHOU, Guyue HU, Xudong GUO, Zezheng ZHANG, Xiaoxiao WEN, Huajun TANG, Xin DONG, Jixiang CHEN, Jinge WEI, and Ningbo CHEN. Their expertise across various disciplines definitely provide our group with more intersting science to explore.

In the end, I would like to express my heartfelt gratitude to the funding support for my research studies in HKU. They are Hui Pun Hing Endowment Fund for Hui Pun Hing Memorial Postgraduate Fellowships (Mainland Collaborations), HKSAR Government Scholarship Fund for Reaching Out Award, University Research Committee for Conference Support for Research Postgraduate Students, Research Grants Council of the Hong Kong Special Administrative Region of China (*HKU* 17205321,

HKU 17200219, *HKU* 17209018, *HKU C*7047 − 16*G*, *CityU T*42 − 103/16 − *N*) and the National Natural Science Foundation of China (*NSFC*) (*N_HKU*712/16).

Huade MAO
The University of Hong Kong
February 20, 2024

List of Publications

JOURNALS:

[1] **Huade Mao**, Yue Yu, Yu-Xuan Ren, Ka Yan Chan, Jiqiang Kang, Xiankai Sun, Edmund Y. Lam, Kenneth K. Y. Wong, "Neural Optimizer for Inverse Design of Complex-Modulated Hologram Implemented by Plasmonic Metasurfaces," *Advanced Photonics Research*, August 2022.

[2] **Huade Mao**, Yu-Xuan Ren, Yue Yu, Zejie Yu, Xiankai Sun, Shuang Zhang, and Kenneth K.Y. Wong, "Broadband meta-converters for multiple Laguerre-Gaussian modes," *Photonics Research*, vol. 9(9), pp. 1689-1698, August 2021.

[3] **Huade Mao**, Pingping Feng, Yu-Xuan Ren, Jiqiang Kang, Kevin K. Tsia, Wa-Tat Yan, and Kenneth K. Y. Wong, "Optical coherence tomography with balanced signal strength across the depth for pearl inspection," *OSA Continuum*, vol. 3(7), pp. 1739-1750, February 2020.

[4] Jiawei Shi, Can Li, **Huade Mao**, Yuxuan Ren, Zhi-Chao Luo, Amir Rosenthal, and Kenneth K. Y. Wong, "Grüneisen-relaxation photoacoustic microscopy at 1.7 µm and its application in lipid imaging," *Optics Letters*, vol. 45(12), pp. 3268-3271, June 2020.

[5] Yu-Xuan Ren, Xinglin Zeng, Lei-Ming Zhou, Cihang Kong, **Huade Mao**, Cheng-Wei Qiu, Kevin K Tsia,and Kenneth K. Y. Wong, "Photonic nanojet mediated backaction of dielectric microparticles," *ACS Photonics*, vol. 7(6), pp. 1483-1490, May 2020.

[6] Yi Zhou, Yu-Xuan Ren, Jiawei Shi, **Huade Mao**, and Kenneth K. Y. Wong, "Buildup and dissociation dynamics of dissipative optical soliton molecules," *Optica*, vol. 7(8), pp. 965-972, August 2020.

[7] Yu-Xuan Ren, Gwinky GK Yip, Lei-Ming Zhou, Cheng-Wei Qiu, Jiawei Shi, Yi Zhou, **Huade Mao**, Kevin K Tsia, and Kenneth K. Y. Wong, "Hysteresis and balance of backaction force on dielectric particles photothermally mediated by photonic nanojet," *Nanophotonics*, vol. 11(18), pp. 4231-4244, August 2022.

CONFERENCES:

[1] **Huade Mao**, Jiqiang Kang, and Kenneth K.Y. Wong, "Inversely design a phase mask for an extended depth of focus through adjoint optimization," in *The 13th International Conference on Metamaterials, Photonic Crystals and Plasmonics*, 2023.

[2] Yu-Xuan Ren, **Huade Mao**, Cihang Kong, Bo Li, and Kenneth K.Y. Wong, "Heat-mediated optical manipulation of Janus particle energized by photonic nanojet," in *The 15th Pacific Rim Conference on Lasers and Electro-Optics (CLEO Pacific Rim, CLEO-PR)*, 2022.

[3] Yu-Xuan Ren, Yi Zhou, **Huade Mao**, Jiawei Shi, Xiaomeng Cui, Kevin K Tsia, and Kenneth K.Y. Wong, "Hysteresis in backaction force mediated by photonic nanojet," in *Frontiers in Optics*, 2020.

[4] Jiawei Shi, Can Li, **Huade Mao**, and Kenneth K.Y. Wong, "All-fiber thulium-doped fiber laser (TDFL) for volumetric photoacoustic microscopy of lipids," in *2020 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*, 2020.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

There are four fundamental forces [1]: strong, weak, electromagnetic, and gravitational forces. When we are talking about optics, it is actually under the frame of electromagnetic force. Nowadays optics has a very rich content to make a subject of its own. But fundamentally, optics is electromagnetic waves, whose behavior is formulated as Maxwell's equations. Here, I want to lay out (1) how electricity and magnetism is connected as electromagnetic waves, which is known as optics in our field; (2) current research directions in optics; (3) literature review of metasurface optics; and (4) my motivation for metasurface research.

1.1 Optics and electromagnetic waves

The connection between optics and electromagnetism is well presented with great detail in literature [2] [3]. I will summarize it below so as to enable a broader range of audience to appreciate optics research in a better position.

There are four findings that lead to the formulation of Maxwell's equations, which is the basis of most solutions to optics problem. Michael Faraday, in 1831, discovered that a changing magnetic field will generate a current [4]. It predicts a time varying magnetic field will produce an electric field. Carl Friedrich Gauss, in 1835, mentioned about the flux theorem [5], where the flux of electric field for any enclosed surface is proportional to the total charges within. In comparison with electric charge, the magnetic monopoles do not exist [2]. By adopting Gauss's Law, the flux of magnetic field in any closed surface is always zero. André-Marie Ampère analyzed the magnetic force between two current-carrying wires [6]. Then the findings were generalized by James Clerk Maxwell stating that a time-varying electric field is the source of a magnetic field [7].

The above four discoveries inspired James Clerk Maxwell to conceptualize the unified theories in electromagnetism [6]. Later, the theory was refined by Oliver Heaviside [8] and demonstrated by Heinrich Hertz [9]. Finally, it evolved into the modern form ubiquitously seen and used as below:

$$\begin{aligned}
\nabla \cdot \mathbf{E} &= \frac{1}{\epsilon_0}\rho \\
\nabla \cdot \mathbf{B} &= 0 \\
\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\
\nabla \times \mathbf{B} &= \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t},
\end{aligned} \tag{1.1}$$

where $\mathbf{E}$ and $\mathbf{B}$ represent electric field and magnetic field respectively, and ϵ_0 and μ_0 denote permittivity and permeability in vacuum, and ρ means the charge density. It depicts the propagation of electromagnetic waves in a well-defined manner.

From Eq. 1.1, it is not hard to obtain the following expressions [3]:

$$\begin{aligned}
\nabla^2 \mathbf{E} &= \epsilon_0 \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \\
\nabla^2 \mathbf{B} &= \epsilon_0 \mu_0 \frac{\partial^2 \mathbf{B}}{\partial t^2},
\end{aligned} \tag{1.2}$$

where ∇^2 is a Laplacian operator. Such an expression is known as differential wave equation with a travelling speed of

$$v = \frac{1}{\sqrt{\epsilon_0 \mu_0}}. \tag{1.3}$$

In International System of Units (SI) [10], permeability μ_0 is assigned a value of $4\pi \times 10^{-7} m \cdot kg/C^2$. Permittivity ϵ_0 can be easily measured from a capacitor [11]. Some simple calculation in Eq. 1.3 is able to predict the travelling speed for electromagnetic waves should be $v \approx 3 \times 10^8 m/s$.

In real life, many experiments were going on to measure the light speed. From celestial observations of Earth–Sun–Jupiter geometry by Olaus Roemer in 1676 [12], to toothed wheel rotating measurement by Armand Fizeau in 1849 [13], and many more others, they all concluded the light speed is roughly $3 \times 10^8 m/s$.

This amazing agreement of speed led Maxwell to believe light is a form of electromagnetic waves [6]. And it also embarked on a new journey to perceive and research optics from a new and fundamental perspective for generations to come.

1.2 Current research overview in optics

While optics and electromagnetism is well explained in Ch. 1.1, it would be easier for audience to appreciate the flourishing research in optics and the position of metasurface

optics, which is my research topic, in the frame of general optics.

I would like to classify those research topics in terms of their scale.

When the size is extremely large, we are referring to astronomy or cosmology research, which ususally involves the development of optical instrumentation [14]. In 17^{th} century, Galileo pointed his telescope towards the sky, where he observed Moon, Jupiter, Venus, and Sun [15]. It is the starting point of us human beings exploring the universe. Nowadays telescopes become amazingly advanced, e.g., Giant Magellan Telescope [16]. To better understand the data from telescopes and observe celestial bodies, it demands more expertise from the researchers, including general relativity [17].

Then, we will dive into the field of geometry optics, where the size should be at least ten times larger than wavelength, but much smaller than celestial distance so as to obviate the complicated usage of general relativity. This field is more closely related with human life. For applications, we have optical imaging [18], optical microscopy [19], optical communication [20], etc. For wavelength, the community spans from ultraviolet [21], visible light [22], infrared [23], to millimeter waves [24]. A recent surge in computer vision research [25] is mostly dealing with problems within this field.

Next, we will explore the field of wave optics. The size should be comparable to the considered wavelength so the geometry optics rule will fail and wave equations shall apply. It solves the problem of interference and diffraction [26]. Usually geometry optics and wave optics are not as separated as how I made the division here, but combined together in a system to tackle a problem. Another huge topic includes nanophotonics [27], ranging from metasurface, plasmonics, to chips, gratings, etc. The knowledge of electrodyamics is a must and finite element method (FEM) is often adopted.

If the size is decreasing to an atomic level, we are entering the field of quantum optics, where light is better to be treated as photons. Lots of interesting research are ongoing, including quantum entanglement communication [28], quantum logic gates [29], parametric oscillation [30], etc. An important branch in biomedical studies involves fluorescence [31], where photon is interacted with molecules with energy absorption or emission.

Lastly, in the subatomic level, the theory is still incomplete. Scientists generally would have to acquire knowledge of quantum field theory. The exploration for quark, electron, muon, and many more, will never end.

1.3 Metasurface optics and my motivation

Metasurface is a subject based largely on geometry optics and wave optics as I classified in Ch. 1.2. Metasurface is a flat integrated with a certain arrangement of nano-antenna, which could arbitrarily manipulate light properties, including its phase, amplitude, spin and orbital angular momentum [32]. The study of optical response of a nano-antenna

usually involves wave optics and electrodynamics, while integrating a metasurface into an optical system may require geometry optics.

Here I will outline the research background of metasurface and my motivation to explore this field.

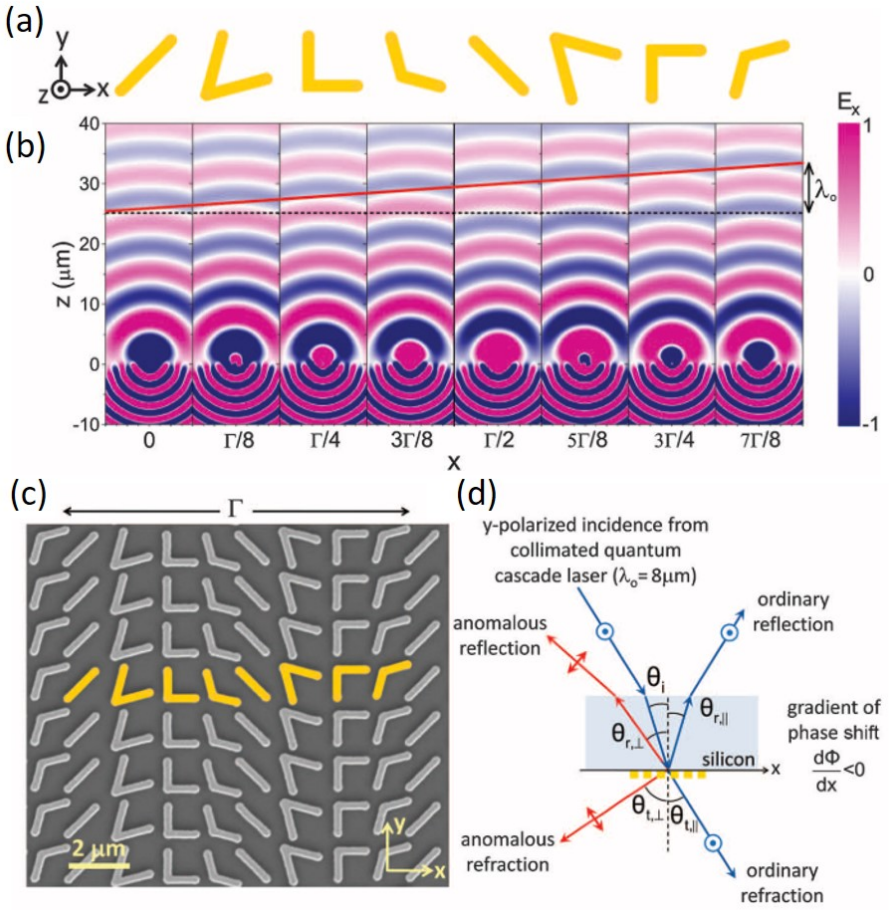

Figure 1.1: Arbitrary reflection and refraction through metallic arrays [33]. (a) The geometry of eight different gold structure. (b) The finite-difference time-domain (FDTD) simulation shows a full-range phase response. (c) Scanning electron miscroscpe (SEM) image of a fabricated device. (d) Schematic figure of arbitrary reflection and refraction.

Metasurface emerged as a popular research topic in optics a decade ago. It consists of periodically distributed subwavelength structures, either metallic or dielectric, to resonance with the incoming electric and magnetic field thus presenting an unexpected electromagnetic response [34]. These certain electromagnetic responses, if well designed and carefully patterned on a surface, will alter optical properties that traditional devices (e.g. a lens or a lens system) cannot achieve. An early example [33] demonstrated anomalous reflection and refraction on a surface composed of metallic arrays in Fig. 1.1. By carefully designing the nano golden structure, they can achieve a phase response from 0 to 2π. According to the generalized Snell's law, the reflection and refraction angle will be altered in will if the phase gradient is introduced appropriately.

It may seem controversial with traditional reflection and refraction law, but fundamentally, it still obeys Fermat's principle, where a deviated propagating direction is caused by artificial phase difference.

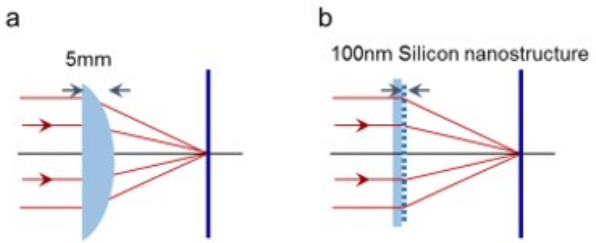

Figure 1.2: Comparison between conventional lens and metalens [35]. (a) A conventional lens focusing a beam. (b) A metalens focusing a beam.

Apart from those intriguing applications, one important feature of metasurface is its miniatured size. When focusing an incoming beam, traditional lens will have a thickness of roughly 5 mm while a metasurface could be as thin as 100 nm [35] in Fig. 1.2. The miniaturization has a huge impact on tons of imaging modalities. For example, a metalens has been proposed to perform near 180° field-of-view (FOV) focusing and imaging [36] within a size of 2 mm in comparison to its bulky fisheye counterpart. In a camera modality, a bifacial metasurface that folds optical path in a cavity was implemented to introduce a 80% reduction in imaging distance with the same imaging performance [37].

Metasurface also comes at another advantage of reduced aberration. Conventional lenses suffer from chromatic aberration [38], astigmatism [39], coma error [40], etc. F. Balli designed a hybrid achromatic metalens [41] through a recursive ray-tracing simulation to achieve a broadband ($1000 - 1800$ nm) focusing efficiency. When it comes to optical coherence tomography (OCT), a metalens was tailored to compensate the phase shift caused by the cylindrical outer sheath within an endoscope [42]. They achieved a near diffraction-limited focusing by correcting the astigmatism. When the incident angle is large to an imaging system, coma aberration will occur. An optimization algorithm was developed [36] to iteratively update the phase profile of the metalens so as to decrease the coma error.

Metasurface can also be tailored to realize various functions. This group designed a metasurface in visible region to manipulate hue-saturation-brightness in a bid for full-colour holography [43]. Another group brought up an idea to control optical amplitude and phase independently at a pixel level [44]. When the pixels were well patterned, a three dimensional (3D) holography was accomplished. Metasurface can be introduced to solve mathematical problems. This group used an electromagnetic field as an input value, implemented a metasurface as an integral solver, and retrieved the solution from the output field [45].

Active tuning is, in addition, an interesting feature for some newly-emerged metasurface technologies. This group, by utilizing a microelectromechanical systems (MEMS),

designed a tunable metasurface doublet. 1-μm change in the distance of two metalense can cause a 36-μm tunability in the focal length [46]. A reconfigurable metasurface was proposed to generate a dynamic hologram when being stretched, paving the way for novel optical display [47]. If a multilayered structure is tuned at optical frequencies, novel phenomena will occur, including pulse compression, spectral manipulation, forbidden frequency gap closure, etc [48].

In comparison, conventional optical solutions have limits. (1) They are bulky. In the case of a camera, multiple lenses have to be aggregated to achieve a desired focus with a minimal aberration. (2) They are sensitive to a broadband spectrum. The optical response at one wavelength may deviate a lot for other wavelengths. Therefore, the overall performance will suffer for a broadband signal. (3) They are not tailored. The performance will deteriorate in a different environment, e.g. biological tissue. (4) They are passive devices. Once manufactured, they are not able to change afterwards.

In conclusion, given metasurfaces' strength in miniaturization, functionality, and tunability, I would like to explore the field in metasurface thus delivering a more powerful optical solution to improve human being's "vision".

1.4 Finite-Difference Time-Domain (FDTD) method

When a photonic device becomes more complicated, it is increasingly impossible to solve the wave propagation property analytically through Maxwell's equation in Eq. 1.1. Then finite-difference time-domain (FDTD) method comes into place. It solves Maxwell's equation on a mesh and computes electric field $\vec{E}$ and magnetic field $\vec{H}$ at every grid points. The software we have used throughout the whole thesis include Lumerical Inc., Meep [49], and Tidy3D.

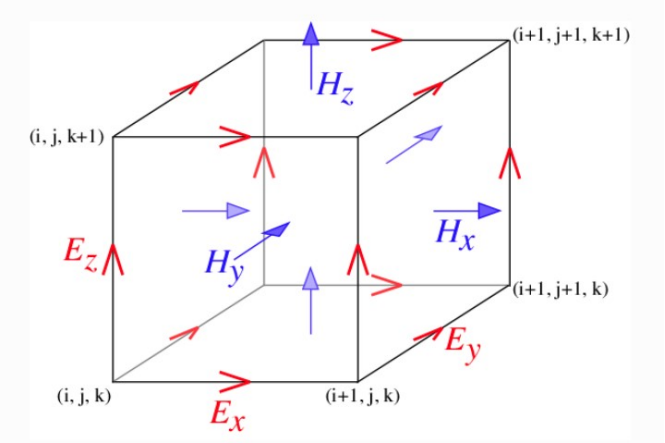

Figure 1.3: Yee lattice to discretize Maxwell's equation in space [49].

FDTD method is based on Yee lattice [50] shown in Fig. 1.3. Firstly, time and space are discretized into a rectangular grid. Then, the electric field $\vec{E}(\vec{r}, t)$ and magnetic field $\vec{H}(\vec{r}, t)$ are solved per time step. And in the end, these electric field and magnetic field

at all grid points are used to calculate other EM properties, e.g. power flux, amplitude, and more.

1.5 Thesis structure

This thesis mainly presents the design and application of metasurface devices to improve "vision" in terms of signal transmission, holographic display, and biomedical imaging. My research work is summarized as follows. (1) The first manufactured device can convert the incoming plane wave into multiple orthogonal modes, which is functional over a 400-nm broadband wavelength range. Since it can expand communication channel, it could transmit more signal to have a better “vision”. (2) The second work utilizes a machine learning approach to inversely design a photonic device to project a desired holographic image. The prediction process is about 400 times faster than theoretical calculation. It could be applied in holographic “vision” scenarios like virtual reality or augmented reality. This work is under review process. (3) The third work is still ongoing with some initial progress. It is about building a tailored fiber to perform vessel imaging through optical coherence tomography. Some nano structures will be etched on the fiber facet to achieve a longer focus of depth. It will have huge potential in biomedical “vision” tasks, e.g. vessel imaging.

In Chapter 2, I explained the design process of a metasurface device to convert a plane wave into multiple Laguerre-Gaussian (LG) modes, thus expanding the communication channels by one more dimension. The LG modes are characterized by radial index and azimuthal index. The azimuthal index is known as orbital angular momentum (OAM). It is used in optical communication due to the orthogonality. So signals won’t interfere during transmission. In comparison, the radial index is largely overlooked since it is not supported by single-mode fibers. However, the radial index could be transmitted through a graded-index fiber or in free space. Therefore, I incorporated both radial index and azimuthal index as the base modes and demonstrated how multiple different modes were integrated on a silica chip and demultiplexed through illumination. This photonic device consists of arrays of gold block unit with a width of 200 nm and length of 200 – 400 nm. When illuminated by a left-handed circularly polarized (LCP) light, the gold block unit will induce a new scattering field upon the incident field. Then, the right-handed polarized (RCP) component of the output field is filtered to achieve a complex modulation with 0 -1 amplitude range and $0 - 2\pi$ phase range. Such complex modulation makes it possible to combine multiple LG modes whereas its phase-only counterparts don’t. The error analysis suggests that this photonic device is robust over a 400 nm wavelength range and experimental demonstration also corroborates.

In Chapter 3, I demonstrated a neural network to accelerate the inverse design of a photonic device, which could generate a desired holographic image upon lensless projection. The design process is about 400 times faster than theoretical computation. The

metasurface in question is capable of complex modulation. Here I developed a convolutional neural network with dual-output, featuring the complex pattern in terms of the amplitude and phase branch. Then the complex pattern, termed as computer generated hologram (CGH) was fabricated into a metasurface and positioned in an experimental setup to project a holographic image. The averaged training loss was 4.43% while the testing loss was 4.57%, suggesting a good generalization. The experimental images, including HKU letters and a Chinese character “guang”, were presented in our submitted manuscripts. I also revealed the neural network from inside to investigate what has been learned through the training. In conclusion, the accelerated inverse design is envisaged to play a role in real-time lensless image projection, e.g. virtual reality, etc.

In Chapter 4, I presented my work on tailoring a fiber facet to build an endoscope which can achieve an extended depth of focus for vessel imaging through optical coherence tomography. Current issues in endoscopic optical imaging include limited penetration depth, optical aberration, trade-off between lateral and transverse resolution, etc. Here I designed subwavelength structures on the beveled fiber facet, which will be etched by focused ion beam (FIB). The design process was accomplished through adjoint optimization. Firstly, a geometrical optics solution has been proposed for benchmark comparison. The structures vary in geometry to reflect the incident light with a phase modulation ranging from 0 to 2π. Then, a wave optics solution will be calculated to obtain the grating structure on a fiber facet, which will produce an electric field response with an extended depth of focus. By applying these two pattern, the reflected beam will be tested of its beam profile. If it can achieve a customized focus depth surrounding the fiber, I will integrate the meta-fiber into an imaging modality. When the beam is transmitted out of the probe and scattered by the tissue and propagate back along the fiber, I will collect the signal and conduct the Fourier analysis to reconstruct the tissue. The tailored fiber endoscope will promote biomedical imaging and disease diagnosis for vessels regarding clinical needs.

In Chapter 5, I concluded my research work regarding the "vision" improvement in three aspects. Multiple orthogonal modes were incorporated on one chip to expand communication band, so as to gain a better signal “vision”. A convolutional neural network was developed to accelerate the inverse design of a metasurface, in the hope of delivering a high-rate holographic “vision”. I am still working on tailoring a fiber for vessel and capillary imaging with a corrected biomedical “vision”.

Chapter 2

Signal "vision": metasurface for signal transmission

Metasurface provides miniaturized devices for integrated optics. Here, I designed and realized a meta-converter to transform a plane-wave beam into multiple Laguerre Gaussian (LG) modes of different orders at various diffraction angles. The metasurface was fabricated with Au nano-antennas, which varies in length and orientation angle for modulation of both the phase and the amplitude of a scattered wave, on a silica substrate. Our error analysis suggests that the metasurface design is robust over a 400-nm wavelength range. This work presents the manipulation of LG beams through controlling both radial and azimuthal orders, which paves the way in expanding the communication channels by one more dimension (i.e., radial order) and demultiplexing different modes.

2.1 Introduction

The LG mode, a solution of the Helmholtz equation in cylindrical coordinates, characterized by the radial index [51], p, and the azimuthal index [52], l, has attracted tremendous attention recently owing to the ability to encode information [53, 54, 55, 56]. The azimuthal index, known as the orbital angular momentum (OAM) [52], has shown applications in object detection [57, 58], optical communication [59], holography imaging [60], etc., mainly counting on the momentum conservation during propagation [61]. Meanwhile, the LG mode, as a complete orthogonal basis [57], has been demonstrated to increase the communication speed [62, 63], e.g., multiplexing and demultiplexing in multiple orthogonal OAM channels [63]. However, the radial index has been largely underexplored by the community as it is not supported by the single-mode fiber used in many optical setups. However, this is not a concern in the free space. Since the radial index is a valid quantum number [64] and could be transmitted through a graded-index fiber [51] or in free space, it has the potential to further increase the capacity of communication system [51] by one more dimension (i.e., radial order).

However, many studies primarily focused on the generation [53, 56, 65] or the observation [66, 67] of LG modes. For example, a mode-sorter [54] was proposed to generate up to 325 LG modes by utilizing 7 phase masks to transform an array of Gaussian beams into the quasi-complete set of LG mode beams. Despite its technical brilliance, it still transforms each Gaussian beam into a single LG mode but not multiple ones, not to mention the setup's complexity.

Here, I report a complex-modulated metasurface to simultaneously tailor multiple LG modes. Many previous optical devices for LG modes generation [54, 56, 68] feature phase-only modulation, which need an iterative algorithm [69] to minimize the error between the output and target field. In recent years, complex modulation [70, 71] was proposed to generate LG modes. However, they were only able to generate one LG mode at a time. By comparison, our metasurface could (1) generate the field in a faster way (no need for iteration) and (2) achieve more sophisticated goals (simultaneous conversion and demultiplexing for multiple LG modes). Our metasurface consists of a 2D array of Au nano-antennas on a glass substrate coated with ITO (indium tin oxide). Our experiments suggest that the designed metasurface performs in a broadband wavelength spanning over 400 nm, which is also corroborated by simulation. As the metasurface is a promising miniaturization technique [33, 72, 73, 74] and many high-order LG modes are achieved through a single chip [56, 75], our work paves the way to future LG modes application and communication channel expansion.

2.2 Theory for complex modulation

The building block of the metasurface is the nano-antenna made of Au [Fig. 2.2(a)], with the height h and width w_x fixed respectively at 80 nm and 200 nm throughout this chapter. The orientation of the nano-antenna ('atom') defines the Pancharatnam-Berry phase modulation [76, 77, 78, 79].

When a left circularly polarized (LCP) light impinges onto this unit, the Au block, functioning as a locally defined birefringent crystal, could alter the amplitude and phase for the orthogonally decomposed light respectively along the fast-axis and slow-axis, thus transforming a certain amount of LCP into right circular polarized (RCP) light [77, 80]. The RCP component [44] from transmitted light is

$$S_{out} = \langle R|\Gamma(-\alpha)\hat{Q}\Gamma(\alpha)|L\rangle \tag{2.1}$$

where R and L denote RCP and LCP respectively; $\Gamma(\alpha)$ is the rotation matrix, α is the orientation angle of the block in Fig. 2.1(a) and $\hat{Q}$ is transformation matrix. The schematic conversion is demonstrated in Fig. 2.1(a).

The complex modulation of the output field is detailed as follows.

I set two basic orthogonal electric field unit as

$$E_x = \overrightarrow{e_x} \cdot exp(i(kz - \omega t)) \tag{2.2}$$

$$E_y = \overrightarrow{e_y} \cdot exp(i(kz - \omega t)) \tag{2.3}$$

where k denotes the wavevector, z the propagation distance along z axis, ω the radius frequency, and t is the time.

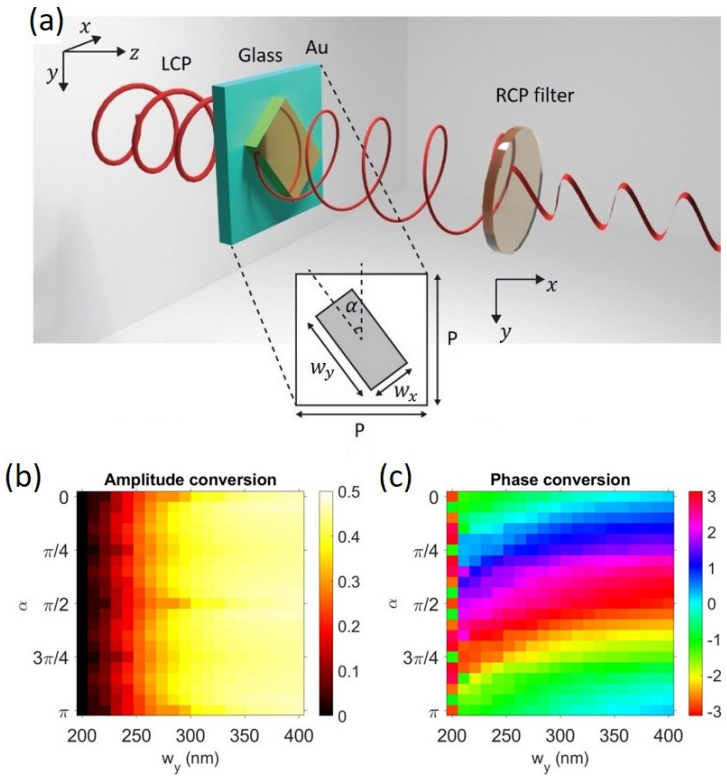

Figure 2.1: The optical response of a nano-antenna. (a) Schematic demonstration of converting LCP to RCP with desired amplitude and phase. The Au block is placed onto a glass substrate. The parameters include orientation angle α, width along x-axis w_x and y-axis w_y. (b) The amplitude conversion and (c) the phase conversion under different w_y and orientation angle α when w_x is set at 200 nm and incident source is set at 1000 nm.

Then the light field could be written in Jones vector form:

$$|x\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{2.4}$$

$$|y\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{2.5}$$

where $|x\rangle$ and $|y\rangle$ denotes the polarized field along x-axis and y-axis, respectively.

As LCP and RCP could be decomposed into two linear polarization, LCP and RCP could be written as follows

$$|L\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix}, \tag{2.6}$$

$$|R\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix}, \tag{2.7}$$

where $|L\rangle$ and $|R\rangle$ denotes LCP and RCP carrying unit power, and i represents the $\pi/2$ phase shift between two linear polarizations.

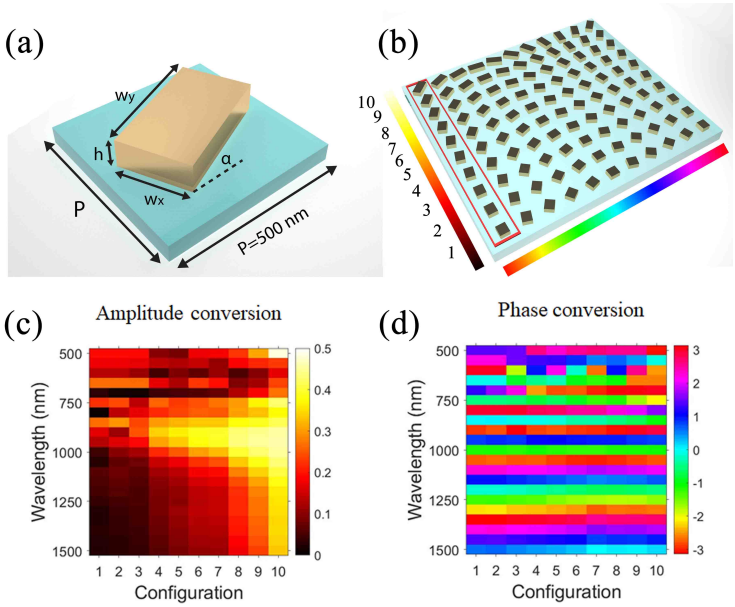

Figure 2.2: Full library of nano-antennas to accomplish complex modulation. (a) Configuration of a unit block. P: period; w_x: width; w_y: length; h: height; α: orientation angle. (b) Configurations of Au block to accomplish complex modulation. The color bars are amplitude range of [0, 0.5] and phase range of [-π, π], which are the same as in (c) and (d). (c, d) Amplitude and phase conversion over 500-1500 nm range for the ten configurations specified in the red rectangle in (b).

In our metasurface, each periodic unit block contains the glass substrate and Au nano-block as in Fig. 2.2(a). The configuration of Au block is determined by 3 parameters, namely, the width w_x, the length w_y, and orientation angle α.

The Au nano-block, considered as a birefringent crystal, has their ordinary and extraordinary refractive index as n_o and n_e along x and y axis given no orientation angle. If the incident light is LCP, output field after Au block's tuning is

$$|\Psi\rangle = R(-\alpha)\hat{Q}R(\alpha)|L\rangle, \tag{2.8}$$

where

$$R(\alpha) = \begin{pmatrix} cos\alpha & -sin\alpha \\ sin\alpha & cos\alpha \end{pmatrix}, \tag{2.9}$$

with α being the orientation angle of the Au block,

$$\hat{Q} = \begin{pmatrix} A_o e^{i\phi_o(z)} & 0 \\ 0 & A_e e^{i\phi_e(z)} \end{pmatrix} \tag{2.10}$$

with A_o and A_e denotes the transmission coefficient along ordinary and extraordinary direction, while $\phi_o(z)$ and $\phi_e(z)$ represents the phase modulation resulting from the birefringence.

In the end, I shall extract the RCP component of the output field $|\Psi(z)\rangle$ [44], which is

$$\begin{aligned} E_{out} &= \langle R|\Psi(z)\rangle \\ &= \frac{1}{\sqrt{2}}(1 \quad -i)^* \cdot |\Psi(z)\rangle \\ &= i \cdot sin(\frac{kd(n_o - n_e)}{2}) e^{i(\frac{kd(n_o+n_e)}{2}+2\alpha)}, \end{aligned} \tag{2.11}$$

where d is the height of our proposed Au block.

From above, I could have the output field specially modulated in amplitude and phase as

$$A_{out} = abs(E_{out}) = sin(\frac{kd(n_o - n_e)}{2}), \tag{2.12}$$

$$\phi_{out} = angle(E_{out}) = \frac{kd(n_o + n_e)}{2} + 2\alpha + \frac{\pi}{2}. \tag{2.13}$$

In phase modulation, the dynamic phase $kd(n_o + n_e)/2$ is determined by the Au block's own parameter, and the Pancharantnam-Berry phase 2α is dependent on orientation angle, which means if the Au block is rotated from 0 to π, the output phase could have a $0 - 2\pi$ modulation range, which is consistent with the phase conversion in Fig. 2.1(c).

The FDTD simulation results are presented in Fig. 2.1(b, c). I set width w_x and height h to 200 nm and 80 nm respectively and alter length w_y from 200 to 400 nm. Such range could reach the maximum amplitude conversion range, e.g., from 0 to 0.464, and is within our fabrication capacity. If w_x equals to w_y, the Au block will show no birefringence. It means no LCP will be converted to RCP, which is consistent with the dark area in Fig. 2.1(b) where w_x is roughly equivalent to w_y.

I selected 10 different configurations with length ranging from 220 nm to 400 nm while keeping the period P of the unit as 500 nm, and orientation from 0 to π to encode the full phase and amplitude of the light wave. In order to offset dynamic phase's difference, each configuration has a unique initial orientation angle as shown in the red block

[Fig. 2.2(b)]. The performance of 10 configurations was evaluated in the wavelength ranging from 500 to 1500 nm using FDTD simulation. The complete conversion map of both the amplitude and phase are revealed in Fig. 2.2(c) and 2.2(d).

2.3 Description of Laguerre-Gaussian(LG) mode

The electric field of a certain Laguerre-Gaussian beam can be written as [81]

$$LG_p^l = \frac{\sqrt{2}r^{|l|}}{w_0^{|l|+1}} \cdot exp(-\frac{r^2}{w_0^2}) \cdot LP_p^{|l|}(\frac{2r^2}{w_0^2}) \cdot exp(il\phi) \tag{2.14}$$

where p is the radial index; l is the azimuthal index that represents the magnitude of OAM; r and ϕ are the cylindrical coordinates; w_0 is the beam waist; LP is the Laguerre polynomial.

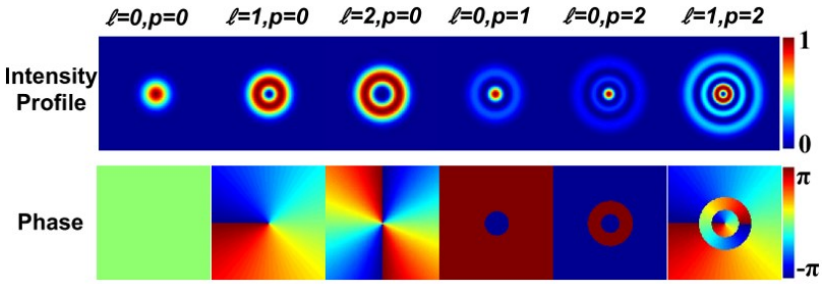

Figure 2.3: Intensity profiles and phase fronts of Laguerre–Gaussian (LG) beams with different azimuthal indices and radial indices [81]. p: radial index. l: azimuthal index.

The intensity and phase profile of some basic LG modes are demonstrated in Fig. 2.3 with varying radial index p and azimuthal index l [81].

Many methods have been proposed to generate one LG mode, which are summarized in Fig. 2.4. A diffractive optical element(DOE) has been demonstrated [82] to transform a Gaussian beam into an LG_3^3 mode beam in Fig. 2.4(a,b). A metasurface[56] can also be used to generate an LG_3^2 mode beam in Fig. 2.4(c,d). Other group fabricated a spiral phase plate with a specific phase mask to deliver an LG_3^1 mode beam, which is shown in Fig. 2.4(d,e,f).

2.4 Metasurface field

The laser beam is converted to LCP before shining on the metasurface under test [Fig. 2.5]. The RCP component of the forward scattering light will carry the information encoded with multiple LG modes deflected into different angles (demultiplexing).

In the past, devices featuring only phase modulation, whether a metasurface [56] or diffraction filter [68], have been proposed for generating LG modes, which is elaborated in Ch. 2.3. However, for more than two LG modes' encoding, phase-only modulation

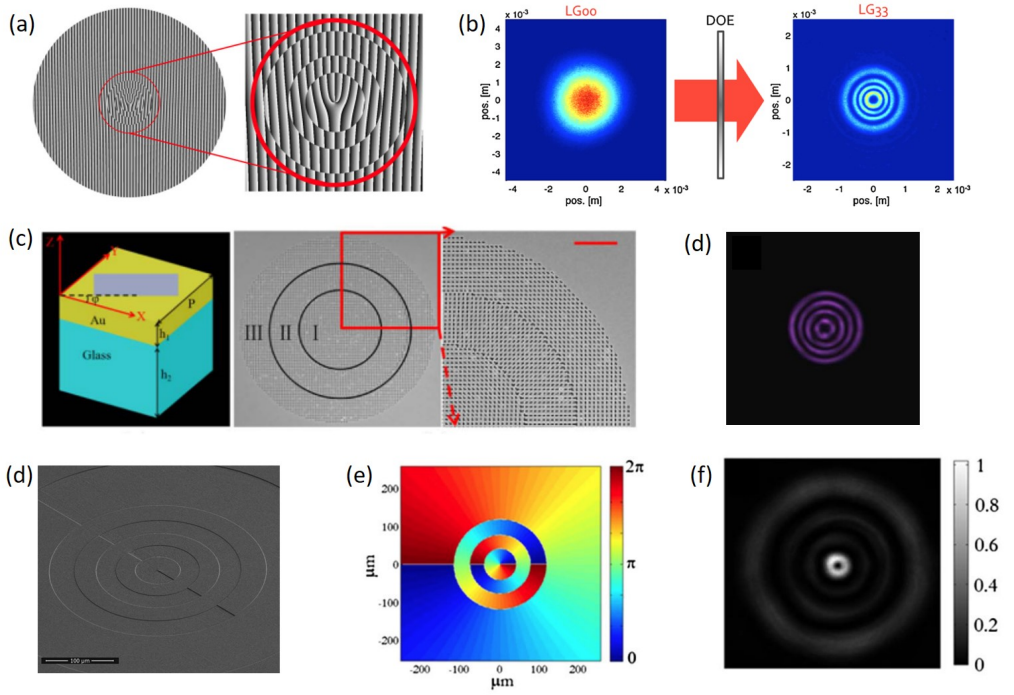

Figure 2.4: Different methods to generate LG modes. (a) A diffractive optical element to transform a Gaussian beam into an LG_3^3 mode beam shown in (b) [82]. (c) A metasurface to generate an LG_3^2 mode beam as in (d) [56]. (d) A spiral phase plate with a phase mask of (e) to deliver an LG_3^1 mode beam in (f) [83].

could not achieve zero-error [detailed in Ch. 2.8]. Complex modulation [70, 71], proposed in recent years, has not been explored in multiple LG modes encoding.

Here, I show that an incident plane-wave beam can be converted into multiple LG modes diffracted into arbitrary directions in transmission by a metasurface, whose complex modulation is expressed as

$$\begin{aligned} M &= \sum_s \sigma_s (LG_p^l)_s \cdot exp(i \cdot sin\theta \cdot \overrightarrow{g_s} \cdot \overrightarrow{\rho}) \\ &= A_M exp(i\phi_M), \end{aligned} \tag{2.15}$$

where σ_s represents different light assigned to the specified LG mode; LG_p^ldenotes the normalized complex distribution of each mode; θ is the deflection angle while g_s and ρ are the deflection direction and pixel vector respectively [inset of Fig. 2.5]. The pixel vector $\vec{\rho}$ points from the center point to one pixel on the metasurface. The amplitude and phase parts of the output field in Eq. (2.15) could be achieved by the metasurface with complex modulation which is fabricated by electron-beam lithography.

The meta-convertor allocates the incident light energy into different decomposition modes according to the designed information magnitude. As an example, I convert the plane-wave beam into two LG modes with non-identical deflection angles. Fig. 2.6(a)

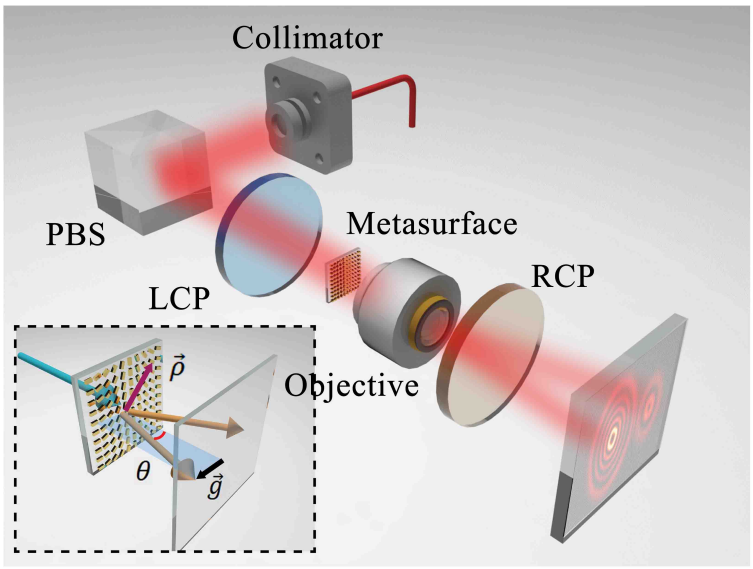

Figure 2.5: Experimental setup. PBS: polarization beam splitter. LCP: LCP generator, consisting of a polarizer and a $\lambda/4$ waveplate. RCP: RCP filter, composed of a polarizer and a $\lambda/4$ waveplate.

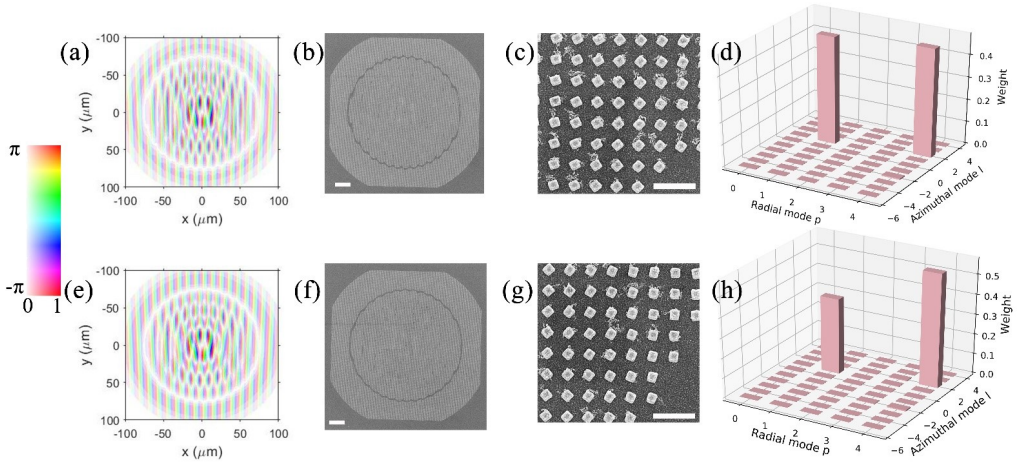

Figure 2.6: SEM figures of metasurface and LG decomposition display. (a) The rounded complex pattern to convert the fundamental Gaussian beam into a combination of LG_4^2 and LG_1^1 directed to $\vec{g_1} = (1, 0)$, with angles $\theta = 5^o$, and $\theta = 10^o$. The assigned weight has a ratio of $LG_4^2 : LG_1^1 = 1 : 1$. (b) Overall scanning electron microscope (SEM) image in (a). (c) Zoomed SEM image in (b). (d) LG decomposition results for the complex pattern in (a) without deflection. (e) The rounded complex pattern to convert the fundamental Gaussian beam into a combination of LG_4^2 and LG_1^1 directed to $\vec{g_1} = (1, 0)$, with angles $\theta = 5^o$, and $\theta = 10^o$. The assigned light has a ratio of $LG_4^2 : LG_1^1 = 6 : 4$. (f) Overall SEM image in (e). (g) Zoomed SEM image in (f). (h) LG decomposition results for the complex pattern in (e) without deflection. Scale bar: 20 μm in (b,e) ; 1 μm in (c,f).

shows the complex modulation of the combination of LG_4^2 and LG_1^1 modes with their deflection direction being respectively $\overrightarrow{g_1} = (1,0)$ and $\overrightarrow{g_2} = (1,0)$, with deflection angle θ of 5^o and 10^o, and an amplitude ratio of 1:1. I fabricated the metasurface by a mapping from the complex field [Fig. 2.6(a)] to the exact configuration in Fig. 2.2(b). Fig. 2.6(b) shows the SEM image of the fabricated metasurface with a zoom-in view shown in Fig. 2.6(c), which elaborates the building blocks of various angles. The metasurface is composed of 400 × 400 unit-blocks. Since each unit-block has a size of 500 × 500 nm, the metasurface has an area of 200 × 200 μm.

I decompose the complex pattern to determine the coefficient for each mode as [detailed in Ch. 2.9]

$$C_p^l = \frac{\langle LG_p^l, U\rangle}{\langle LG_p^l, LG_p^l\rangle} \tag{2.16}$$

where U denotes the complex field without deflection to maintain central symmetric and the denominator is the normalization term. This process is one kind of Fourier decomposition with the Fourier basis being the LG modes which form the complete set [84]. The result [Fig. 2.6(d)] suggests a ratio of 1:1 between the modes of LG_4^2 and LG_1^1, which is well consistent with our design.

Here, I analyzed the effect of fabrication error, detailed in Ch. 2.7. Under the assumption, ±10 nm deviation of width and length of the Au-block will cause an average error of mostly lower than 10%, which is acceptable. Then, I also evaluate the influence of fabrication error on the decomposition results [detailed in Ch. 2.9]. The decomposition would have no significant change provided that the fabrication error has a mean of 0. The decomposition error deviates only 4.5% for a mean of 0.05, which suggests the robust tolerance to the fabrication error. This has also been validated with another design [Fig. 2.6(h)] with a combination of $LG_4^2 : LG_1^1 = 6 : 4$.

2.5 Metasurface fabrication

The metasurface was fabricated by electron-beam lithography (EBL).

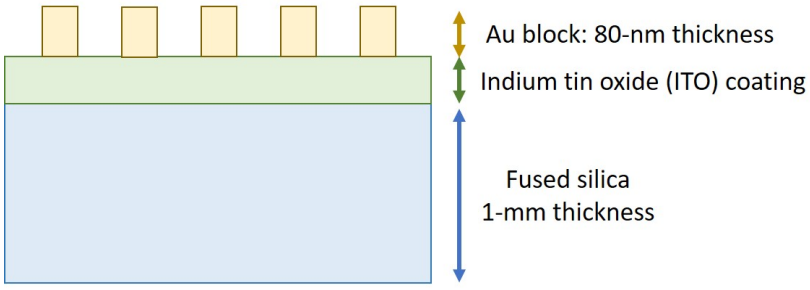

Figure 2.7: The cross section of the fabricated metasurface. The substrate is fused silica with a thickness of 1 mm, which has an indium tin oxide (ITO) coating. The nano-antenna consists of gold with a fixed thickness of 80 nm.

The cross-section of the fabricated metasurface is shown in Fig. 2.7. The substrate is fused silica with a thickness of 1 mm, which has an indium tin oxide (ITO) coating. The nano-antenna consists of gold with a fixed thickness of 80 nm. The size of the substrate is $25 \times 25\ mm^2$, which is ordered from UniversityWafer. The size of every nano-antenna is demonstrated in Fig. 2.2(a). The period P, height h, and width w_x are respectively fixed at 500 nm, 80 nm, and 200 nm. The length w_y of each nano-antenna ranges from 200 to 400 nm with different orientation angles. The length and orientation angle can be specified from Table. 2.1 at a determined amplitude and phase modulation.

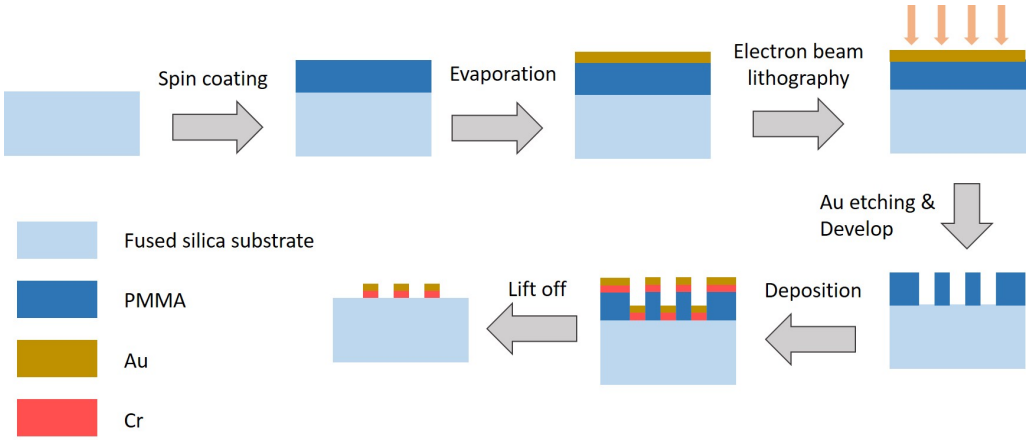

Figure 2.8: Electron-beam lithography (EBL) fabrication process. At first, the glass substrate was spin coated with a layer of poly methyl methacrylate (PMMA), followed by electron beam(e-beam) evaporation of thick Au overcoat.Then the sample was exposed in Elionix ELS7800 EBL system. Next, the Au layer was removed by Au etchant solution, followed by rinsing with deionized water and finally blow drying with nitrogen. Exposed samples were developed in a mixture of water and isopropyl alcohol. Next, Cr was deposited by e-beam evaporation followed by Au with a thickness of 80 nm. After the deposition, the PMMA resist and metal atop were removed via lift-off in acetone to reveal the metasurface.

The EBL process is demonstrated in Fig. 2.8. We collaborated with Prof.Sun Xiankai, Dr.Yu Zejie, and Dr.Yu Yue from The Chinese University of Hong Kong (CUHK) in terms of EBL fabrication.

The standard EBL process is well explained in great detail in many literatures [85][86]. Here I will only brief the operation flow of EBL process. At first, the glass substrate was spin coated with a layer of poly methyl methacrylate (PMMA), followed by electron beam(e-beam) evaporation of thick Au overcoat to prevent the charging effect during EBL. The graphic data system(GDS) file of the desired metasurface was uploaded to the EBL system. The sample was exposed in Elionix ELS7800 EBL system at a certain certain accelerating voltage and beam current. After the e-beam exposure, the Au layer was removed by Au etchant solution, followed by rinsing with deionized water and finally blow drying with nitrogen. Exposed samples were developed in a mixture of water and isopropyl alcohol. Next, Cr was deposited by e-beam evaporation

followed by Au with a thickness of 80 nm. After the deposition, the PMMA resist and metal atop were removed via lift-off in acetone to reveal the metasurface.

The imperfection of the fabricated metasurface can be traced back to the EBL machine Elionix ELS7800 in CUHK. During our collaboration period, the accelerating voltage of the machine was unstable.

2.6 Simulation and experiment

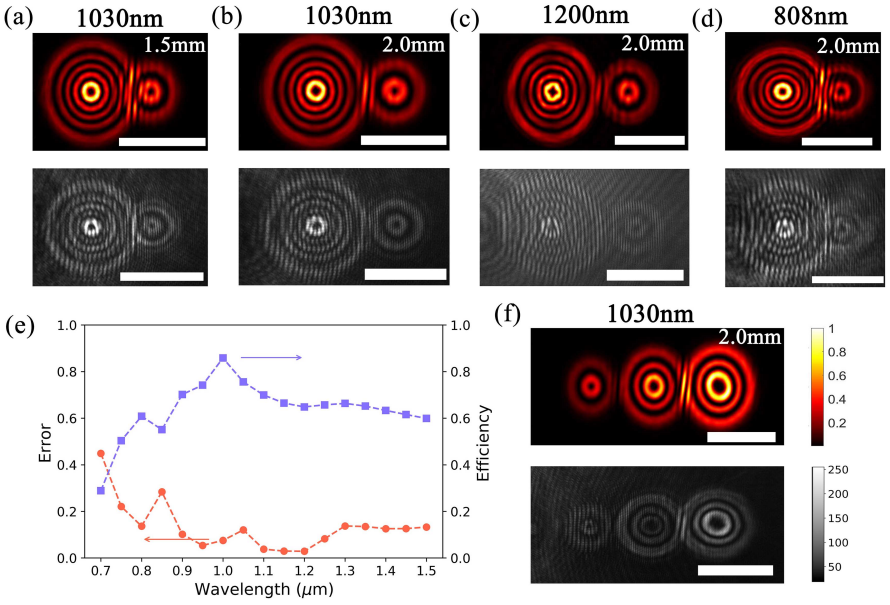

Figure 2.9: Experimental results: Broadband performance of meta-converters. (a-d) Diffraction patterns with metasurface designed for $LG_4^2 : LG_1^1 = 6 : 4$. The test wavelength and the measurement distance from the metasurface are all labeled with the beam profile. First row: simulation. Second row: experiments. (e) Simulated broadband performance for the meta-converter of $LG_4^2 : LG_1^1 = 6 : 4$ in terms of error and efficiency. (f) Metasurface contains modes of $LG_1^1 : LG_2^2 : LG_2^3 = 3 : 4 : 5$. The left edge corresponds to zero deflection for all figures. Scale bar for all: 200 μm.

The holographic image of the metasurface could be calculated through a complex transmission function [44, 87],

$$H(x,y) = \sum_{m,n} M_{mn} \cdot \frac{exp(ikR_{mn}(x,y))}{R_{mn}(x,y)} \tag{2.17}$$

where M_{mn} represents a pixel on the metasurface determined by Eq. (2.17); k is the wavenumber; $R_{mn}(x,y)$ denotes the distance between M_{mn} and the position (x,y) on the holographic image H. The final normalized intensity is calculated as

$$I_{norm} = \frac{\langle H, H \rangle}{\langle H, H \rangle_{max}}. \tag{2.18}$$

The experimental setup is demonstrated in Fig. 2.5. The PBS is to convert the incoming beam into a polarized beam. The LCP is to convert a polarized beam into an LCP beam, while the RCP is to filter out the RCP component of the transmitted beam. The lasers we used are at three separate wavelengths, namely, 808 nm, 1030 nm, and 1200 nm. The CCD is from IDS imaging with an item number of UI-1240SE-M-GL. LCP is a polymer circular polarizer from Thorlabs (CP1L1064). RCP is also a polymer circular polarizer from Thorlabs (CP1R1064). The selection criteria is that they are readily available from our lab or Thorlabs.

I next evaluated the broadband performance of the metasurface. The complex pattern M_{mn} of a certain wavelength was converted using the broadband conversion map for 10 configurations [Fig. 2.2(c, d)]. The metasurfaces were designed at the wavelength of 1000 nm. The complex pattern varies slightly from 500 nm to 1500 nm [Supplementary visualization in [88]]. Then I altered the wavenumber k in Eq. (2.17) to simulate the holograhic image.

The supplementary visualization [88] shows how the complex pattern and the LG decomposition results change over the broadband wavelength range. The visualization 1 is about the metasurface featuring modes of $LG_4^2 : LG_1^1 = 6 : 4$. In the video, the first row contains the amplitude and phase map of the complex field without deflection over the wavelength range from 500 nm to 1500 nm. The third figure of the first row is the bar chart of the LG decomposition results. The second row contains the complex field of the metasurface with deflection, which is the fabricated complex field I use for experiments.

Experimentally, the performance of the fabricated metasurface is evaluated at various wavelengths (1030 nm, 1200 nm, and 808 nm) to demonstrate the broadband modulation ability. The diffractive images are collected at a distance of 1.5 mm or 2.0 mm from the metasurface [Fig. 2.9]. Fig. 2.9(a) shows the diffraction pattern for λ = 1030 nm. The LG_4^2 and LG_1^1 deflects to roughly 130 μm and 260 μm at a propagation distance of 1.5 mm, which correspond to the diffraction angles of 5^o and 10^o, respectively. As seen from both the hologram simulation and the experiment [Fig. 2.9(a)], there are apparent interference fringes in the region where the two LG modes overlap, which is owing to the insufficient separation in the near field. However, the fringes will disappear when the paired beams propagate further (2 mm away from the metasurface), as shown in Fig. 2.9(b).

In Fig. 2.9(c, d), I tested the performance of the meta-converter at two other wavelengths of 1200 nm and 808 nm, and the experiments (bottom) agree well with the simulation results (top). I analyzed its error and efficiency from 700 nm to 1500 nm [Fig. 2.9(e)] through simulation. The decomposition error measures the deviation of energy allocation from design and the efficiency measures the ratio of energy flowing into desired modes defined in Eq. (2.35) [detailed in Ch. 2.9]. Take the metasurface in Fig. 2.6(f) for example. The desired modes and their amplitude ratio is $LG_4^2 : LG_1^1 = 6 : 4$. After normalization, the desired coefficients vector is $\overrightarrow{\omega} = (0.6, 0.4)$. After decomposition at

1 μm, the decomposed vector is $\overrightarrow{\sigma} = (0.504, 0.354)$. The error is calculated as 0.076 and the efficiency is 0.858, which are both acceptable. I repeat the procedure from 700 to 1500 nm with a step of 50 nm to get Fig. 2.9(e) . It shows that from 800 nm and up to 1500 nm, the metasurface has an error about or below 10% mostly (except 28% at 850 nm). Although the above analysis adopts two LG modes, in general, the design and analysis procedure could apply to a combination of multiple LG modes. For demonstration, I show the three LG modes encoding with an energy ratio of $LG_1^1 : LG_2^2 : LG_2^3 = 3 : 4 : 5$. Since the RCP filter is not at 100% efficiency, the unconverted LCP passing the metasurface will form an interference pattern resembling its arrangement, which will cause a concentric ring interference at the center in Fig. 2.9(a-d).

Although the result is decent for three LG modes here, the metasurface has an issue of energy transmission efficiency. The maximum amplitude modulation is 0.464. The transmitted power is 21.5% at most. Given the metasurface is complex modulated, the whole energy transmission would be even lower. If I increase the amount of LG modes, there must exist a limit where the energy allocated to each mode is undetectable in experiment. Future work about dielectric metasurface to improve the energy transmission is prospected.

More experiment and simulation results are demonstrated in Fig. 2.10. Two metasurfaces are fabricated from Fig. 2.6(a,e). In Fig. 2.10(a,c,e), I introduce a first metasurface containing mode LG_4^2 and LG_1^1 with their deflection direction being respectively $\vec{g_1} = (1, 0)$ and $\vec{g_2} = (1, 0)$, with deflection angle θ of 5° and 10°, respectively, and a weight ratio 1:1 between them. It was tested under three wavelengths, namely, they are 1030 nm, 1200 nm, and 785 nm. The second metasurface in Fig. 2.10(b,d,f) is designed to generate a combined mode of LG_4^2 and LG_1^1 with the same deflection direction and angle as the first sample, but a different ratio of 6:4. The testing wavelengths were the same with the first metasurface. The experiments and simulation matches with each other well.

The metasurface that delivers the testing results in Fig. 2.9(a-d) and Fig. 2.10 contains mode LG_4^2 and LG_1^1 with a ratio of $6 : 4$. The SEM figure of this metasurface is shown in Fig. 2.6(b-c). Second metasurface that produces the results in in Fig. 2.9(f) features mode LG_1^1, LG_2^2, and LG_2^3 with a ratio of $3 : 4 : 5$. The SEM figure of second metasurface is demonstrated in Fig. 2.11.

2.7 Fabrication error analysis

Due to the fabrication limit, the maximum amplitude conversion I could obtain is 0.464. Therefore, a projection mapping from 0-1 to 0-0.464 is adopted to yield quasi $0 - 1$ amplitude modulation. And the approximation mentioned above is taken to round the desired amplitude to the nearest one decimal. For example, if the desired amplitude conversion is 0.34, it will be rounded to 0.3 denoted as η_N, which after approximation could be written as $\eta_P = \eta_N \times 0.464 = 0.1392$. Next, I searched over the simulated

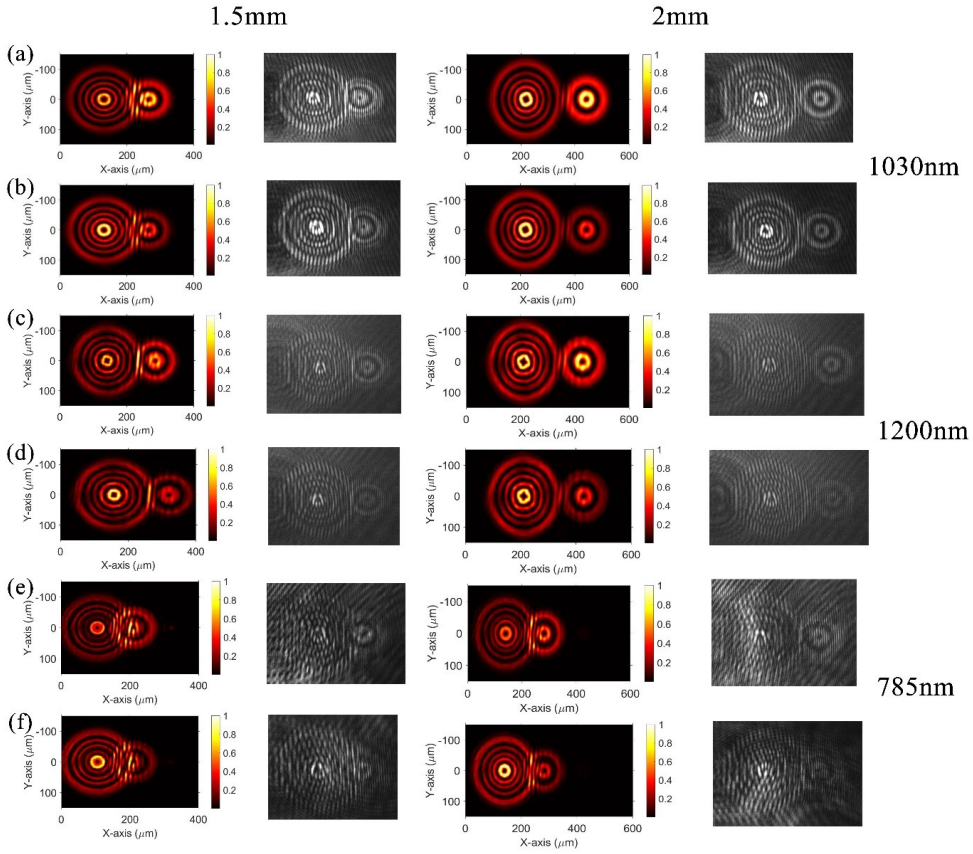

Figure 2.10: More broadband simulation and experimental results. (a, c, e) Metasurface 1 designed in Fig. 2.6(a). (b, d, f) Metasurface 2 designed in Fig. 2.6(e). First and third columns are simulation results based on 1000 nm, 1200 nm, and 800 nm respectively. Second and fourth column are experimental results based on 1030 nm, 1200nm, and 785 nm. I changed the focal distance from 1.5 mm to 2.0 mm to fully separate the two LG modes as captioned in the figure head.

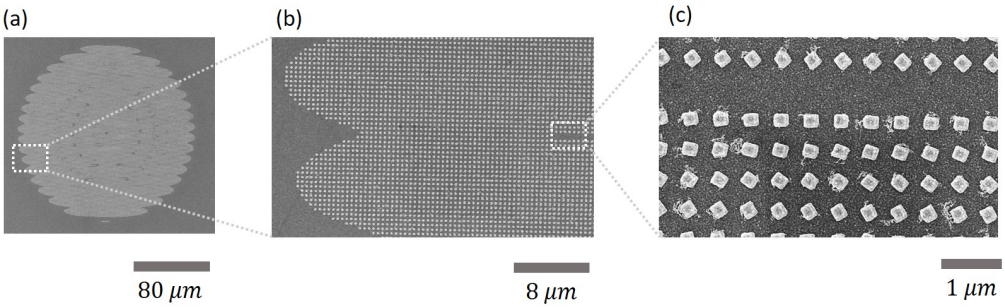

Figure 2.11: SEM figure of the metasurface featuring $LG_1^1 : LG_2^2 : LG_2^2 = 3 : 4 : 5$, whose testing result is demonstrated in Fig. 2.9(f).

conversion map in Fig. 2.1(b) and 2.1(c) to find the appropriate parameters, including the length w_y and initial orientation α_0, on which the recomputed actual amplitude η_{ac}

and phase ϕ_{ac} are based. Last, the relative error ξ is calculated as

$$\xi = \frac{|\eta_P e^{i\times(-1.129)} - \eta_{ac} e^{i\times\phi_{ac}}|}{0.464}, \tag{2.19}$$

where -1.129 is the calibrated phase all configurations need to match at their initial orientation. In this example, after the actual amplitude conversion is computed as 0.1392, the length of the Au block w_y is determined as 230 nm. Then, the initial orientation α_0 is set at $-\frac{1}{60}\pi$, in order to offset the dynamic phase, which in other words is to make the initial phase conversion ϕ_{ac}=-1.177 as close to calibration -1.129 as possible. Under this configuration, I conducted FDTD method to calculate the actual amplitude conversion η_{ac} and error ξ as detailed in Table. 2.1.

The tolerance of the size of the nano-block is critical for the broadband performance. The metasurface is fabricated with the electron-beam lithography. But there is no statistical data reflecting the fabrication error for our metasurface. Empirically, the fabrication has an error of 10 nm. So here I assume the machine will have a 10-nm deviation on both the width and length of the Au-block and the deviation is Gaussian distributed. Therefore, the probability density function of the distribution could be represented as follows

$$p(Z;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{di}{2}}|\Sigma|^{\frac{1}{2}}} \cdot exp(-\frac{1}{2}(Z-\mu)^T\Sigma^{-1}(Z-\mu)), \tag{2.20}$$

where $Z = (w_x, w_y)^T$ is the vector of the length and width of Au-block; $\mu = (\mu_x, \mu_y)^T$ are the mean of the length and width, which are set as the 10 configurations in Table. 2.1; $\Sigma = (\sigma_1^2 \quad 0; 0 \quad \sigma_2^2)$ is the covariance matrix of the Gaussian distribution, where two variables are uncorrelated under our assumption; and di denotes the dimension of the variable.

The total possible range of Z here, in our consideration, is [190, 210] nm for width and $[\mu_y - 10, \mu_y + 10]$ nm for length. I conduct full-wave FDTD simulation over this region, and get the error map

$$\epsilon(Z) = \frac{|C(w_x, w_y; \alpha_0) - C(\mu_x, \mu_y; \alpha_0)|}{0.464}, \tag{2.21}$$

where $C(x, y, ; \alpha)$ denotes the complex modulation under the configuration x and y given orientation α.

The average error ξ could be calculated as

$$\xi = \int_{\mu_x-10}^{\mu_x+10}\int_{\mu_y-10}^{\mu_y+10} \epsilon(Z) \cdot p(Z;\mu,\Sigma) dw_x dw_y, \tag{2.22}$$

where the integral region is the total possible range of Z.

Since the two variables, width w_x and length w_y are uncorrelated. I just draw one variable at a time in Fig. 2.12. If the standard deviation σ_i (i=1,2) is small enough,

Table 2.1: Parameters Approximation. η_N: desired quasi-amplitude conversion. η_P: desired actual amplitude conversion. w_y: length of the nano-antenna along y-axis. α_0: initial orientation angle. Given the predetermined length w_y and orientation α_0, the actual amplitude and phase conversion are denoted as η_{ac} and ϕ_{ac}. ξ: error between the actual and desired complex conversion. Error under three different Gaussian distributions: ξ_1: pessimistic; ξ_2: neutral; ξ_3: optimistic.

η_N	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
η_P	0	0.0464	0.0928	0.1392	0.1856	0.232	0.2784	0.3248	0.3712	0.4176	0.464
$w_y(nm)$	0	220	225	230	250	255	265	270	300	330	400
α_0	0	0	0	$-\frac{1}{60}\pi$	$-\frac{1}{18}\pi$	$-\frac{1}{12}\pi$	$-\frac{1}{9}\pi$	$-\frac{5}{36}\pi$	$-\frac{1}{6}\pi$	$-\frac{7}{36}\pi$	$-\frac{41}{180}\pi$
η_{ac}	0	0.0452	0.103	0.1248	0.181	0.236	0.275	0.326	0.378	0.424	0.45
ϕ_{ac}	0	-1.129	-1.112	-1.177	-1.08	-1.154	-1.208	-1.212	-1.151	-1.112	-1.10
ξ	0	0.258%	2.23%	3.39%	2.17%	1.53%	4.77%	5.82%	2.30%	2.07%	2.90%
ξ_1	0	11.17%	9.29%	8.47%	10.68%	8.39%	9.85%	8.20%	8.62%	6.83%	5.18%
ξ_2	0	9.48%	7.87%	7.28%	9.95%	6.91%	9.01%	7.00%	7.92%	6.30%	4.50%
ξ_3	0	7.49%	6.04%	6.10%	9.38%	5.58%	8.23%	5.89%	7.19%	5.75%	3.96%

Table 2.2: Three scenarios of Gaussian distribution.

std	coverage	level
$\sigma_1 = \sigma_2 = 10$	46.5%	pessimistic
$\sigma_1 = \sigma_2 = 5$	91.0%	neutral
$\sigma_1 = \sigma_2 = 3.33$	99.4%	optimistic

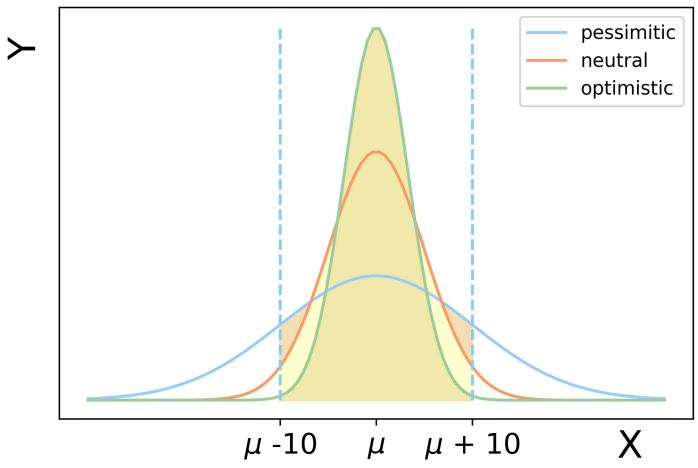

Figure 2.12: Gaussian distribution under three scenarios: "pessimistic", "neutral", "optimistic".

the most distribution will be inside our considering region. Here I assign three levels to different standard deviation, ranging from "pessimistic", "neutral", to "optimistic" in Fig. 2.12 and Table 2.2. When the standard deviation σ is set to 3.33 nm, nearly 99.4% of the whole distribution, which is the yellow area in Fig. 2.12, is within our

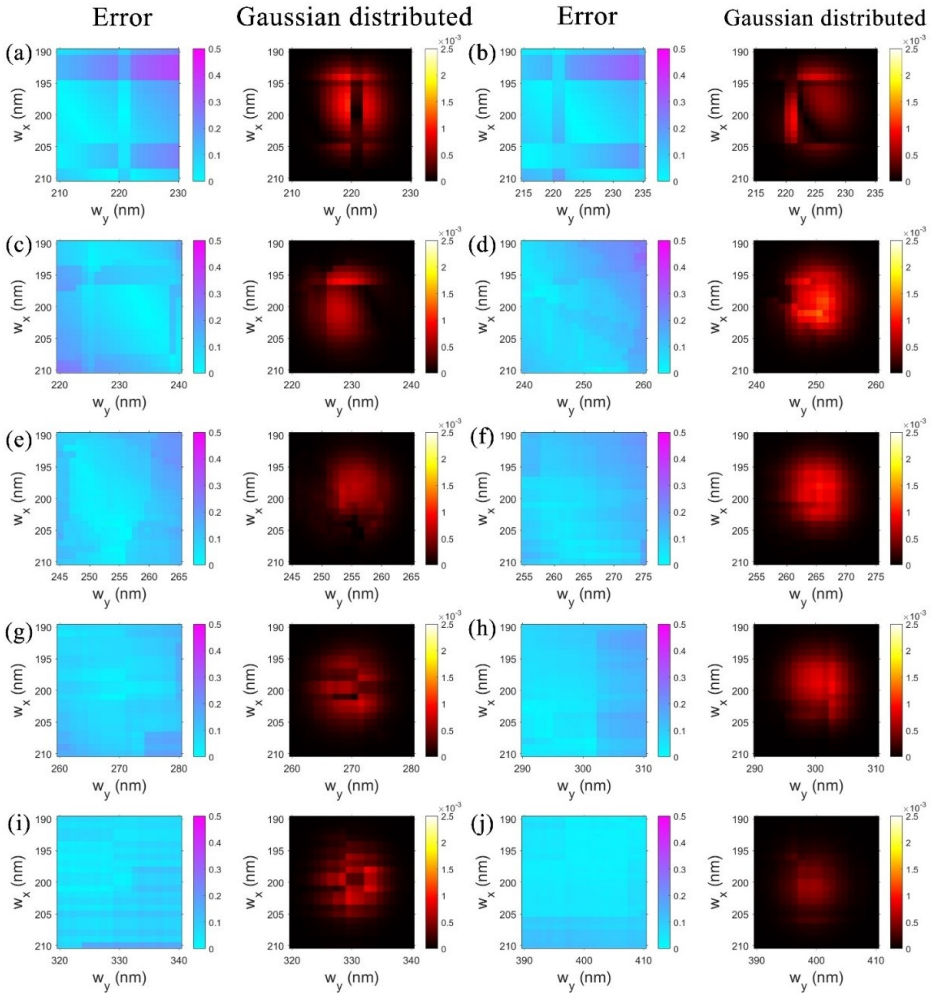

Figure 2.13: Error distribution for 10 different configurations with 10-nm deviation along the width and length of the Au-block. The first and third columns are the absolute error over different width and length. The second and fourth columns are the Gaussian-distributed error from the desired configuration. For the ten configurations I selected here, the width w_x is fixed at 200 nm. The length w_y are (a) 220 nm, (b) 225 nm, (c) 230 nm, (d) 250 nm, (e) 255 nm, (f) 265 nm, (g) 270 nm, (i) 300 nm, (j) 330 nm, and (k) 400 nm.

considering region. When the standard deviation σ goes up to 10 nm, only 46.5% of the distribution falls into our assumption region, which is termed "pessimistic" in this case. I will calculate the total error according to Eq. (2.22) under these three scenarios.

The error map is revealed in Fig. 2.13 for ten nano-block configurations. The first and third column, captioned "Error" in the figure, is the absolute error among the fabrication region. The second and fourth column show the error with Gaussian distribution under "optimistic" assumption.

The average error ξ, elaborated in Eq. (2.22), is calculated under all the three scenarios and listed in Table. 2.1. Even under pessimistic assumption, the average error is about or lower than 10%, which is acceptable. And in Ch. 2.9, I will conclude that the deviation will be fully waived under LG mode decomposition, which shows the robustness of my design.

2.8 Defects of phase-only modulation

Phase-only modulation could not achieve zero-error. The electric field of a LG mode is detailed in Eq. (2.14). Here I simplify it as

$$LG_p^l = A_{pl}(x,y) \cdot exp(i\phi_{pl}(x,y)), \tag{2.23}$$

where A_{pl} and ϕ_{pl} represents the amplitude and phase respectively.

The ideal complex pattern containing exactly the desired LG modes could be written as

$$U = \sum_{p=0}^{\infty} \sum_{l=-\infty}^{\infty} \sigma_{pl} \cdot A_{pl}(x,y) \cdot exp(i\phi_{pl}(x,y)), \tag{2.24}$$

where σ_{pl} is a complex number.

If the metasurface could feature phase-only modulation, the complex pattern could be rewritten as $U = C \cdot exp(i(\phi(x,y))$, where C is a constant. The power is $P = |U|^2 = |C|^2$. Naturally, I will have

$$\begin{aligned} &\sum_{p=0}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{p'=0}^{\infty} \sum_{l'=-\infty}^{\infty} \sigma_{pl}\sigma_{p'l'} A_{pl} A_{p'l'} \cdot \\ &exp(i(\phi_{pl}(x,y) - \phi_{p'l'}(x,y))) = |C|^2. \end{aligned} \tag{2.25}$$

The above equation is not true for any arbitrary combination of more than 2 LG modes. Because given the unimodular property [89], the Fourier basis, here is the LG mode basis, should be either one or infinite in Eq. (2.25). For finite LG modes, which are more than 2 LG modes but not infinite, Eq. (2.25) is impossible. This is quite similar to Ref. [90], where the phase-only modulation for OAM combination is proved impossible.

2.9 LG mode decomposition

I use LG_p^l to denote the Laguerre-Gaussian mode with radial index p and azimuthal index l. The detailed expression is detailed in Eq. (2.14). The field of LG mode LG_p^l is normalized throughout the manuscript, which means the highest amplitude is 1 for all the different modes.

Consider an electric field U, which could be decomposed into the combination of different LG modes as follows

$$U = \sum_{p,l} C_p^l \cdot LG_p^l, \tag{2.26}$$

where C_p^l, a complex number, is the coefficient of the mode LG_p^l, which could be calculated as

$$C_p^l = \frac{\langle LG_p^l, U \rangle}{\langle LG_p^l, LG_p^l \rangle} = \frac{\int {LG_p^l}^* \cdot U dS}{\int {LG_p^l}^* \cdot LG_p^l dS}, \tag{2.27}$$

where ${LG_p^l}^*$ is the conjugate of mode LG_p^l, and the integral area is the whole electric field S. The electric field U is normalized to ensure the highest amplitude is 1. Considering the different LG modes are also normalized, I have $\sum abs(C_p^l) = 1$.

Since LG modes are orthogonal to each other, obviously we have

$$\frac{\int {LG_p^l}^* \cdot U dS}{\int {LG_p^l}^* \cdot LG_p^l dS} = \begin{cases} 1, p = p' \; and \; l = l' \\ 0, otherwise \end{cases} \tag{2.28}$$

Here I only consider the LG modes combination without deflection. It could give a basic idea of how the fabrication error will affect the final LG modes. However, if I introduce deflection into the decomposition, the results will cover a wide range of LG modes basis. It is similar to Ref. [91], where the OAM state is fixed but located away from the center, leading to a non-pure singular OAM in the spectrum. Such simplification could help us quantify the error and determine how much the error will influence our results.

For our metasurface M, detailed in Eq. (2.15), I firstly introduce the complex field U, which is equivalent with M but without deflection.

For the field U, the decomposition is

$$(C_p^l)_k = \frac{\int (LG_p^l)_k^* \cdot U dS}{\int (LG_p^l)_k^* \cdot (LG_p^l)_k dS} = \sigma_k, \tag{2.29}$$

where σ_k represents the weight assigned to mode $(LG_p^l)_k$.

The relative intensity of each mode is written as

$$\begin{aligned} (I_p^l)_k &= \frac{\sigma_k^2}{\langle U, U \rangle} = \frac{\sigma_k^2}{\sum_i \langle \sigma_i (LG_p^l)_i, \sigma_i (LG_p^l)_i \rangle} \\ &= \frac{\sigma_k^2}{\sum_i \sigma_i^2} \end{aligned} \tag{2.30}$$

where the sum of the intensity of all the modes are $\sum_k (I_p^l)_k = 1$. Due to conjugation, the phase information disappears in the intensity calculation step.

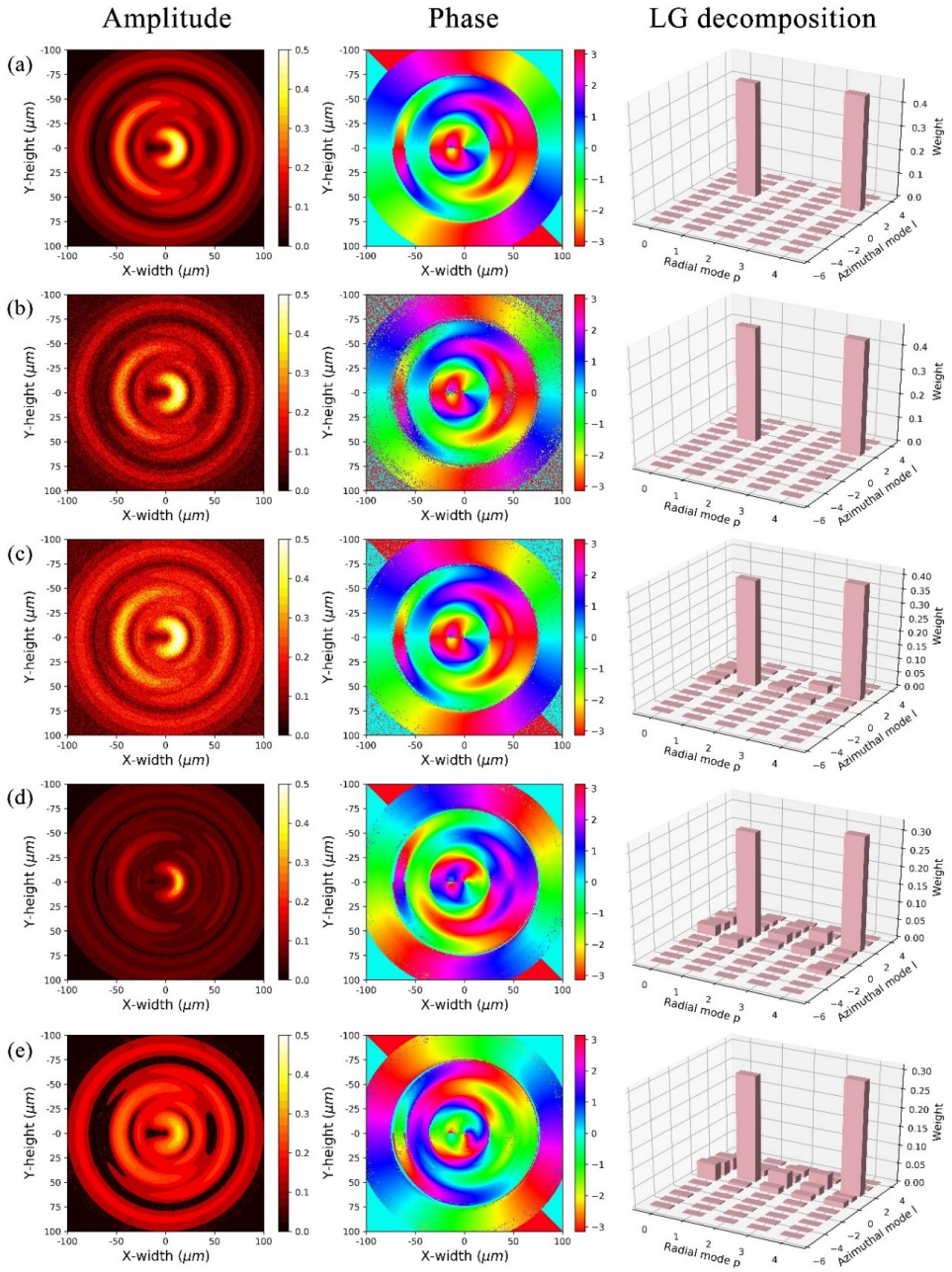

Figure 2.14: LG mode decomposition demonstration. First column: amplitude pattern. Second column: phase pattern. Third column: LG decomposition results. (a) Rounded complex pattern, featuring $LG_4^2 : LG_1^1 = 1 : 1$. (b) Complex pattern with a Gaussian noise $N(0, 0.05^2)$ applied to the amplitude. (c) Complex pattern with a Gaussian noise $N(0.05, 0.05^2)$ applied to the amplitude. (d) Complex pattern without noise for incident light at 1200 nm. (e) Complex pattern without noise for incident light at 800 nm.

From Eq. (2.26), I simplify the electric field as

$$U = A(x,y) \cdot exp(i\phi(x,y)) = A_U \cdot exp(i\phi_U). \tag{2.31}$$

Given the assumption that the fabrication error of the width and length of Au-block are Gaussian distributed, it is reasonable to assume that it will lead to amplitude error being Gaussian distributed, which means the electric field of the fabricated metasurface is

$$U_{fab} = (A(x,y) + t) \cdot exp(i\phi(x,y)), \tag{2.32}$$

where t is the amplitude of Gaussian distribution $N(\mu_a, \sigma_a^2)$. Note the mean μ_a and variance σ_a^2 of the amplitude error here is different from the mean μ and variance σ^2 of fabrication error in Ch. 2.7.

Fig. 2.14 reveals the complex pattern and LG decomposition results with regard to a metasurface, whose complex field is $U = LG_4^2 + LG_1^1$, with the weight ratio being $LG_4^2 : LG_1^1 = 1 : 1$ but without deflection compared to the metasurface field. The decomposition result is also shown in Fig. 2.6(d). Apparently, the error has little effect on our LG mode decomposition.

The deviation of width and length of Au-block will affect the amplitude and phase output. The error in this regard has been obtained through FDTD in Fig. (2.13). I assume the orientation doesn't change, so the Pancharatnam-Berry phase remains the same. The width and length's deviation will affect the dynamic phase change, which is negligible if only considering $\pm 10nm$'s change. Therefore, the width and length's deviation will only affect amplitude. Given the fabrication machine is unbiased, it is intuitive to take the mean μ_a as 0. Under optimistic distribution in Table. 2.1, the expectation of error ξ_3 ranges from 3.96% to 9.38%. It would be reasonable to choose a standard deviation σ_a in this scale. In Fig. 2.14, I assign σ_a a value of 0.05. The Gaussian noise $N(\mu_a, \sigma_a^2)$ is applied to the amplitude part of the combined LG mode pattern.

For demonstration, I write the LG mode as $LG_p^l = A_p^l \cdot exp(i\phi_p^l)$, where A_p^l and ϕ_p^l are both real numbers. The decomposition for the electric field after fabrication is

$$\begin{aligned}
(C_p^l)_{fab} &= \frac{\langle LG_p^l, U_{fab} \rangle}{\langle LG_p^l, LG_p^l \rangle} \rightarrow \int {LG_p^l}^* \cdot U_{fab} dS, \\
&= \int A_p^l (A_U + t) \cdot exp(i(\phi_U - \phi_p^l)) dS \\
&= \int A_p^l A_U \cdot exp(i(\phi_U - \phi_p^l)) dS \\
&\quad + \int t \cdot A_p^l \cdot exp(i(\phi_U - \phi_p^l)) dS.
\end{aligned} \tag{2.33}$$

Since the normalization factor $\langle LG_p^l, LG_p^l \rangle$ is constant, I do not show it explicitly in Eq. (2.33) for simplicity.

Given t is a random variable, I take the expectation to get the final coefficient

$$\begin{aligned}(C_p^l)_{fab} &\leftarrow E[(C_p^l)_{fab}] \\ &= \int A_p^l A_U \cdot exp(i(\phi_U - \phi_p^l))dS \\ &\quad + \int t \cdot N(\mu_a, \sigma_a^2)dt \int A_p^l \cdot exp(i(\phi_U - \phi_p^l))dS \\ &= \int A_p^l A_U \cdot exp(i(\phi_U - \phi_p^l))dS \\ &\quad + \mu_a \cdot \int A_p^l \cdot exp(i(\phi_U - \phi_p^l))dS \\ &= C_p^l + \mu_a (C_p^l)_{phase'}\end{aligned} \tag{2.34}$$

where $(C_p^l)_{phase}$ is the decomposition coefficient of the pure-phase modulation. From Eq. (2.34), I could conclude that as long as the mean of amplitude error is 0, the noise term, which is the second term in Eq. (2.34), could be waived. Even if the amplitude error has some bias, which means μ_a is not 0, the decomposition coefficient $(C_p^l)_{fab}$ will be somewhere between complex modulation C_p^l and phase modulation $(C_p^l)_{phase}$. If the bias is infinitely large (which is quite impossible in the sense), which means the amplitude error is large, the coefficient $(C_p^l)_{fab}$ will degenerate into phase modulation $(C_p^l)_{phase}$.

In order to quantify the deviation caused by the fabrication noise, here I set a criteria to judge the error in LG mode expansion.

The "error" in LG mode expansion here is defined as the sum of unwanted modes and the deviation from desired modes. Our desired coefficients vector of different LG modes is denoted as $\overrightarrow{\omega} = (\omega_0, ..., \omega_i, ..., \omega_n)$, while the coefficients after decomposition is $\overrightarrow{\sigma} = (\sigma_0, ..., \sigma_i, ..., \sigma_n)$, where ω_i and σ_i means the desired and decomposed coefficient for i^{th} mode respectively, and n means all the modes desired.

The error could be calculated as

$$\epsilon = (\sum_i |\frac{\omega_i}{\omega_0} - \frac{\sigma_i}{\sigma_0}|). \tag{2.35}$$

The error could be interpreted as the deviation from the desired modes. I check whether the ratio of different modes is close to our predetermined ratio.

Here I define another term called efficiency, which is

$$\gamma = \sum_i \sigma_i. \tag{2.36}$$

The efficiency is the sum of our desired modes. If a large ratio of poIr is split into unwanted modes, the metasurface will not generate desired modes even if the error ϵ is low.

2.10 Broadband performance

This section shows the metasurface performance over a broader spectrum from 500 nm to 1500 nm. Firstly, I will map our complex pattern at 1000 nm to other wavelength according to the conversion map in Fig. 2.2(c, d). For each wavelength, I conducted LG mode decomposition.

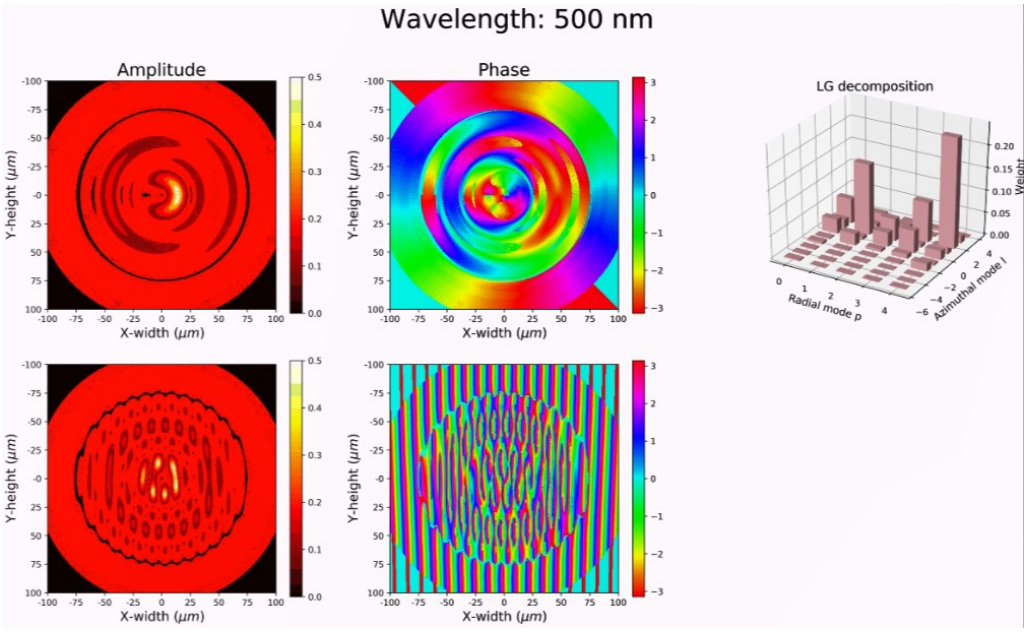

Figure 2.15: The decomposition results at 500 nm. The complex field of metasurface features $LG_4^2 : LG_1^1 = 1 : 1$.

In Fig. 2.15, it shows how the complex pattern of metasurface 1 [Fig. 2.6(a)] and the LG decomposition results at 500 nm incidence. The change over the broadband wavelength range was recorded in a video, which is available for research purposes from the author upon reasonable request.

In Fig. 2.16, it shows how the complex pattern of metasurface 2 [Fig. 2.6(e)] and the LG decomposition results at 500 nm incidence. The change over the broadband wavelength range was recorded in a video, which is available for research purposes from the author upon reasonable request.

For the wavelength from 800 nm and onwards till 1500 nm, the error is about below 10% mostly and the efficiency is above 60%, which is acceptable and validates our metasurface design under a broadband usage. I also notice the anomaly point at 600 nm in Fig. 2.17. The reason is as follows.

I draw the complex conversion for the 10 configurations I selected in Fig. 2.18 at three wavelength, namely, 600 nm, 1000 nm, and 1200 nm. From the LG decomposition in Ch. 2.9, the key to maintain a low error when wavelength shifts is the amplitude is a linear mapping and the phase has only constant shift. And the linear coefficient of the

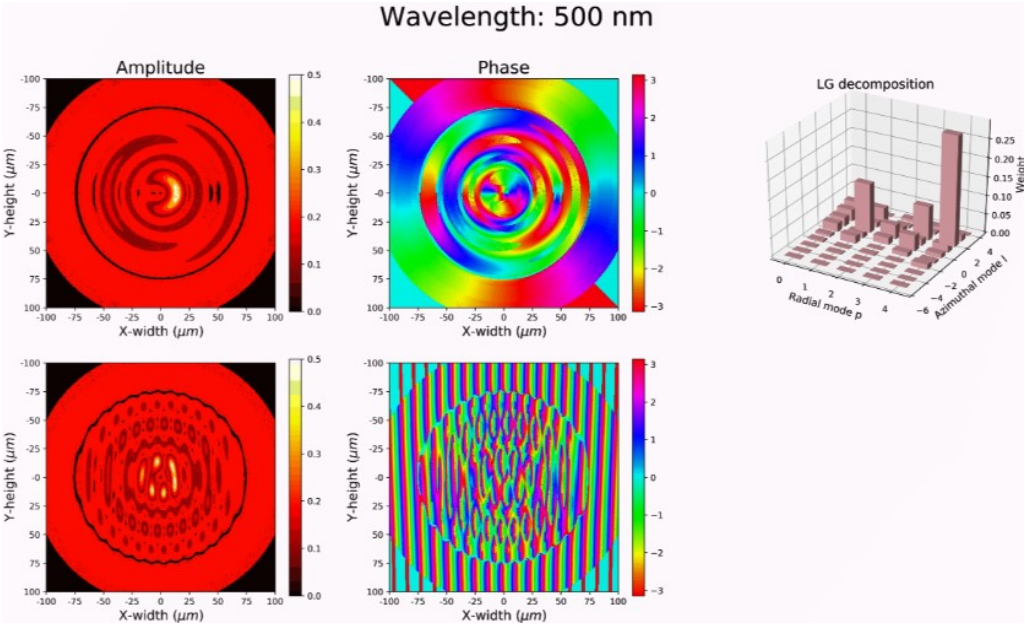

Figure 2.16: The decomposition results at 500 nm. The complex field of metasurface features $LG_4^2 : LG_1^1 = 6 : 4$.

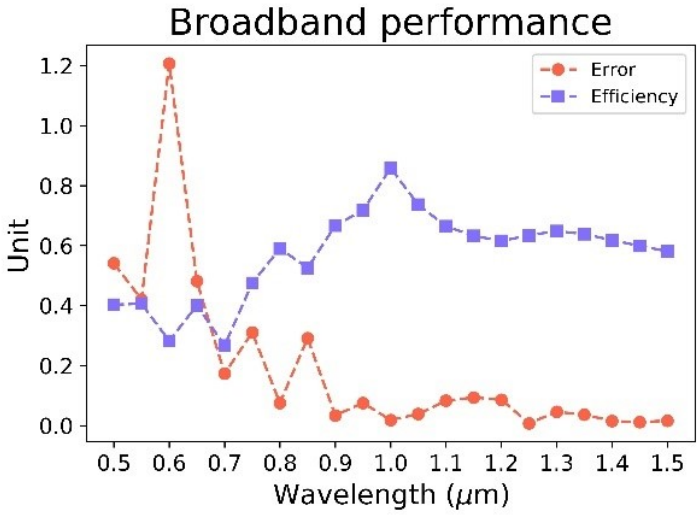

Figure 2.17: . The error and efficiency of the metasurface over the wavelength 500 to 1500 nm. The complex field features $LG_4^2 : LG_1^1 = 1 : 1$.

mapping determines the efficiency. In Fig. 2.18(a), I can see the amplitude conversion at 1200 nm is roughly that at 1000 nm times a constant, termed as linear coefficient. After decomposition and normalization, the results won't deviate much. However, for 600 nm, it has a nonlinear relationship with 1000 nm, which causes the LG information disappear. In Fig. 2.18(b), the phase shift of 1000 nm and 1200 nm is about a constant. In LG decomposition, it will add a phase term in the final LG mode coefficient, which will be eliminated after conjugation. However, for 600 nm, the phase modulation has a

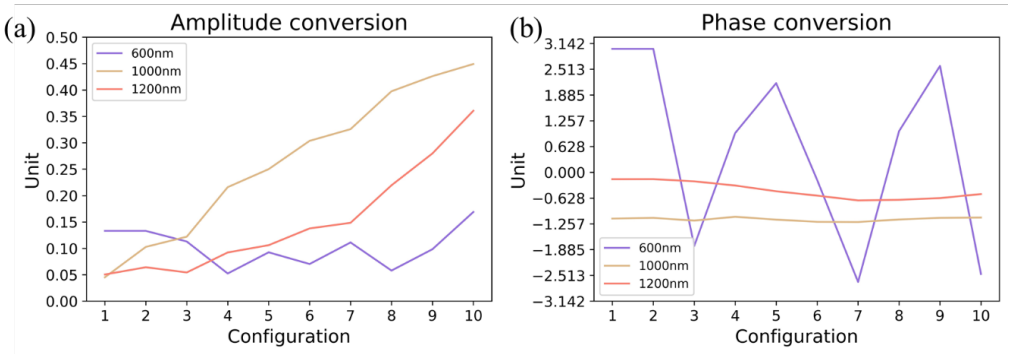

Figure 2.18: (a) Amplitude conversion and (b) phase conversion for wavelength at 600 nm, 1000 nm and 1200 nm, which are obtained from the conversion in Fig. 1(c) and (d) in manuscript.

high variance, which also leads to the LG information's loss. For the metasurface with LG modes of $LG_4^2 : LG_1^1 = 6 : 4$, I conducted the FDTD simulation from 550 nm to 650 nm in a finer grid. The video is available from the author upon reasonable request. The error and efficiency is demonstrated in Fig. 2.19.

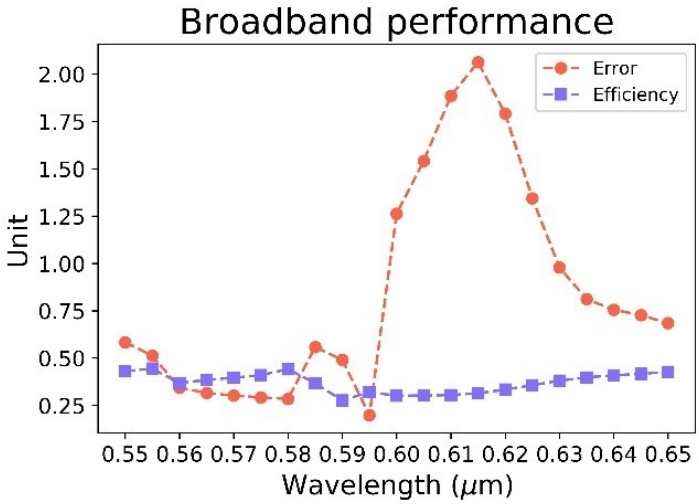

Figure 2.19: Broadband performance for metasurface of $LG_4^2 : LG_1^1 = 6 : 4$. The wavelength range is set from 550 nm to 650 nm.

From Fig. 2.19, the error is high in the region of 600 to 630 nm, which means the metasurface will not have an acceptable performance in that wavelength range.

2.11 Discussion

Here, I demonstrated a metal metasurface to generate multiple LG modes with different diffraction angles. It has a robust performance across a broadband wavelength range from 800 nm to 1200 nm. It has a potential application to increase the communication speed.

For signal transmission, traditional optical communication methods usually adopt SMF, which only allows TE or TM mode beam to transmit. It will limit the communication speed with a maximum bottleneck of allowed orthogonal channels. However, by utilizing LG modes, we will be able to increase the communication orthogonal channels by one more dimension.

It is obvious that many engineering problem will appear if we want to build an LG-mode based communication system. In order to achieve LG mode transmission in a fiber, it requires a micro-engineered fiber to receive and transmit LG modes. We will have to leverage the price of a tailored fiber and the benefit of an improved signal transmission speed.

2.12 Summary

In sum, I designed a metal metasurface to generate multiple LG modes and separate them simultaneously. The design technique was based on forward calculation. The metasurface features with a complex modulation and have a decent performance over 400 nm wavelength range. Although I demonstrate the energy allocation into a limited number of LG modes, the meta-convertor could in principle distribute the light energy into an arbitrary number of modes. Since the radial index of LG mode could be transmitted through a graded-index fiber [51, 92] or in free space, it is envisaged it could potentially increase the communication bandwidth in the optical communication systems with integrated optical devices.

Chapter 3

Holographic "vision": neural network for metasurface design

Last chapter demonstrates the forward calculation method to design a metasurface for signal transmission. However, forward calculation only suffices to solve the case where the mapping from a metasurface to a desired far-field radiation is known. In many cases, for a desired far-field radiation, the metasurface pattern is unknown.

Therefore, this chapter presents a neural network to inversely design a metasurface for arbitrary holographic display. Inverse design of a metasurface involves searching parameters in a high-dimensional space, which needs huge computational power. To ease the computational burden, neural network, a well-researched computer science stream, has demonstrated its potential usage in the inverse design of a photonic device. Many studies primarily focused on the nanostructure's configuration. However, the near field of a metasurface requires further optimization to achieve a desired holographic image at relieved computational power. Here, a convolutional neural network is developed to optimize the complex field of a computer-generated hologram, which can be fabricated into a metasurface to generate the desired holographic image upon lensless image projection. The neural optimizer can accelerate the design speed 400 times faster than theoretical computation and reduce the time complexity from $O(n^4)$ to $O(n^2)$. The neural optimizer has been compared against three other methods, e.g. gradient-based optimization, global genetic algorithm and coupled-mode theory, to demonstrate a lowered error rate from more than 10% to 1.38% for the benchmark testing and a reduced running time from hours to near 1 second. The neural optimizer is envisioned to play a key role in lensless image projection and real-time metasurface pattern design.

3.1 Introduction

Metasurface, as an integrated photonic device, has been used in many fields, including hologram [44][79], polarization control [93], color manipulation [43], achromatic focusing [94], etc. To achieve those goals, the metasurface needs to be meticulously designed

to realize specific optical responses concerning both the nano-antenna configuration and the pattern arrangement . The design process usually consists of sweepings over different parameters, e.g. shape, height, length, etc., through an electrical-magnetic (EM) solver (e.g., Lumerical FDTD) [44] so as to find a suitable configuration.

Given the computation burden in the sweeping process, the deep learning model prevails in finalizing the nano-unit's geometry in a much faster way [95][96]. Apart from nano-antennas' design, the deep learning model has also been applied to optimize the arrangement of them [97]. A generative model is adopted to create a meta-grating with the best diffraction efficiency [98]. In addition, a convolutional neural network is trained to map from the holographic image to the phase mask [99]. Such inverse design problems could be solved readily through a deep learning model [98][100][101][102]. However, with metasurfaces taking on different forms, e.g. gratings, plasmonics, layers, etc., for different applications, it requires a tailored neural network to accommodate the various parameters.

In the field of hologram, the inverse problem from a holographic image to the phase mask is generally solved through Gerchberg-Saxton algorithm [103] or inverse Fourier transform [104], which could be easily achieved through a physical Fourier transform lens [103]. The underlying physics is about Fourier optics. Given a field transmission function $f(x,y)$, we could get its amplitude spectrum in Fourier plane as $F(u,v) = \mathbb{F}(f(x,y))$, where $\mathbb{F}$ stands for Fourier transform [104]. Experimentally it is common to use a spatial light modulator (SLM) to tune the transmission function $f(x,y)$ in a phase-only manner, thus generating the desired holographic image. The phase mask on SLM is usually termed as computer generated hologram (CGH) [105]. Metasurface featuring Pancharatnam-Berry phase could also be used for hologram with higher resolution and better performance [106].

However, the phase-only modulation is apparently far worse than complex modulation in terms of sharpness and luminosity [44]. Moreover, complex modulation is free from the Fourier transform lens, as it incorporates the inverse propagation function into the complex design [87], thus every point in the holographic image (object field) is the interference results of all the spherical wave from the CGH. This lensless modality, combined with metasurface, is of high interest in recent studies [44][99][107]. However, efficient inverse mapping from a desired holographic image to the original complex-modulated metasurface is still highly desirable in the community.

In hologram applications, metasurface also surpasses conventional optics in terms of resolution. Bulky optical elements are usually adopted by conventional holograms to shape the wavefront via phase accumulation effect [108]. However, it will suffer from issues of low-resolution imaging and high-order diffraction. Since metasurface hologram can modulate amplitude and phase at a subwavelength scale, it can achieve high-resolution holographic images without diffraction orders by interacting with the incident wave [109].

Here, to the best of our knowledge, we firstly presented a neural optimizer (NO) to design a complex-field CGH, which was 400 times faster than theoretical computation, to achieve a desired holographic image. The time complexity was reduced from $O(n^4)$ to $O(n^2)$, where n is the length of the CGH. The CGH was then fabricated into a metasurface within a lensless modality to deliver a compact optical system. The NO featured a dual-branch convolutional neural network to encode the complex pattern containing amplitude and phase. A similar architecture has been demonstrated in a classification work [110] to improve the accuracy. In our work, the NO tended to generalize from the holographic image field to the metasurface field. Our NO has outperformed three other methods, e.g. gradient-based optimization, global genetic algorithm and coupled-mode theory, in terms of the running time and error rate in metasurface deisgn. It is envisaged that such technique can promote lensless image projection and real-time metasurface pattern design.

3.2 Theory for computer generated hologram

Holography was first invented by Denis Gabor in 1947 [111]. It utilizes interference and diffraction theory to record and reconstruct the 3D information of a physical object [112].

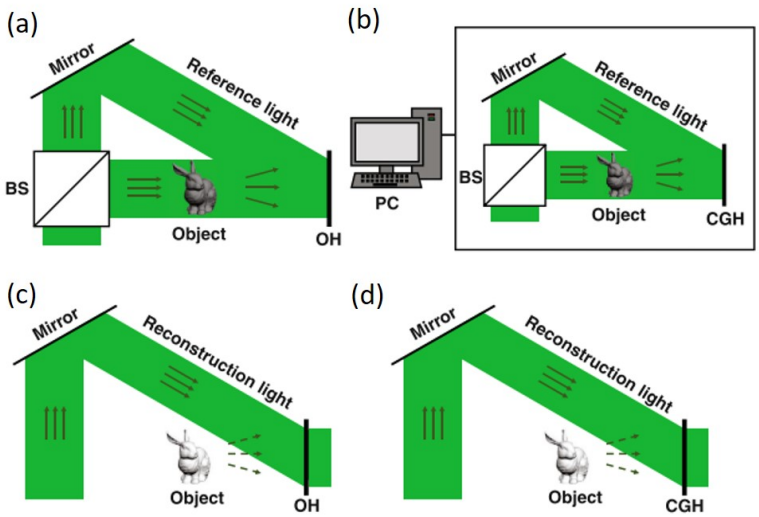

Figure 3.1: The scheme of hologram recording and reconstruction process [112]. (a, c) Optical holography (OH). (b, d) Computer-generated holography (CGH).

For a traditional optical hologram (OH) shown in Fig. 3.1(a,c), the hologram is recorded as the interference fringe pattern of the reference beam and the object beam. Then the original object can be reconstructed when the reconstruction beam is chosen as the reference beam. However, OH can only allow static images and it will pose strict requirements on the stability of optical systems, which limits the application in dynamic display.

Given the boom of computer and optoelectronic technology, computer-generated holography (CGH), shown in Fig. 3.1(b,d), can achieve a more dynamic display with a simpler optical system compared to OH. Some refreshable hologram techniques, e.g. spatial light modulator (SLM), can be used to digitally generate the hologram.

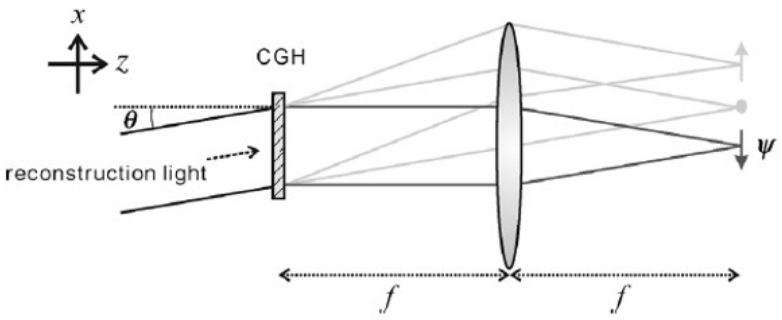

Figure 3.2: Reconstruction scheme for the CGH [113].

A Fourier-type hologram of CGH is demonstrated in Fig. 3.2. It shows the reconstruction of a Fourier hologram through a Fourier transform lens with a focal length f. The relationship between the hologram $H(x,y)$ and the reconstructed image $\varphi(x,y)$ is [113]

$$\varphi(x,y) = \mathbb{F}\{H(x,y)\varphi_r(x,y)\}, \tag{3.1}$$

where $\mathbb{F}$ denotes the Fourier transform, $\varphi_r(x,y)$ represents the complex field of the incoming reconstruction beam, and $\varphi(x,y)$ is the complex amplitude of the reconstructed image.

3.3 Principle for inversely designing a metasurface

This lensless holography could be simplified as the superposition of the spherical wave propagation from each point source on a metasurface. For a spherical wave, the free space propagation can be represented by the Green function [87]

$$G_0(\overrightarrow{r}, \overrightarrow{r_1}) = exp(\pm\frac{ik|\overrightarrow{r} - \overrightarrow{r_1}|}{4\pi|\overrightarrow{r} - \overrightarrow{r_1}|}), \tag{3.2}$$

where the plus (minus) sign denotes the forward (backward) propagation.

Within a metasurface, each unit performs as a secondary source according to the Huygens–Fresnel law [107]. The holographic image is, therefore, the superimposition of the secondary wave from all the units in the metasurface. Inversely, for a given holographic image, a metasurface's pattern is the summation of all contributions of back-propagation from the pixels, which is formulated as

$$M(x,y) = \sum_{i,j} \frac{exp(-ik \cdot R_{ij}(x,y))}{R_{ij}(x,y))} \cdot H(i,j), \tag{3.3}$$

where $M(x,y)$ denotes the complex modulation at position (x,y) on a metasurface M, k represents the wavenumber, H(i,j) is the intensity information at position (i,j) for a given holographic image H, and $R_{ij}(x,y)$ is the pixel-wise distance between M and H, which in mathematical form is $|\overrightarrow{OM}(x,y) - \overrightarrow{OH}(i,j)|$ with O being the coordinate origin.

Equation 3.3 is essentially an inverse design process. However, it involves, apparently, two loops to finalize the complex information of a metasurface. If the holographic image and metasurface are both in size of n×n pixels, the whole inverse design problem is of the complexity $O(n^4)$. Here, through a neural network, it is anticipated to generalize Equation 3.3 into

$$M(x,y) = \hat{Q} \cdot H(x,y), \tag{3.4}$$

where the decoding process $\hat{Q}$ is the transformation matrix which functions as NO. Hence, the problem is optimized to a complexity of $O(n^2)$. Through this process, we managed to accelerate the inverse design 400 times faster compared with the theoretical method in Equation 3.3 with a testing loss below 5%. It has the potential for lensless image projection on a faster basis.

In a coherent imaging system where wavefronts interfere with eath other, speckle will usually occur [114]. In some cases, the community want the speckle to be reduced. Numerous methods have been reported to reduce speckles by introducing angle diversity, polarization diversity, and wavelength diversity [115]. The working principle of those methods is that they have introduced different independent speckle patterns, which can be averaged out on the detector.

3.4 Neural network architecture

Our NO aims at mapping a holographic image to the metasurface's complex pattern. It consists of U-Net [116] and residual network [117] architectures [Fig. 3.3]. Since the U-Net is commonly used in image segmentation and encoding [118], here we extended the layer by adding a short-term memory within each block [Fig. 3.3(f)]. The NO model features two outputs to represent both the amplitude and phase of the complex field of a metasurface. The rationale behind this architecture is (1) it is required to represent a complex number by two real numbers, therefore dual-output is adopted; (2) amplitude and phase are correlated to some extent, hence their skip connections share the same down sample block.

The model, similar to a prior report [99] but different in output branches, comprises five different blocks. Each block color represents (a) down sampling, (b) dilation, (c) up sampling, (d) residual and (e) convolution, respectively [Fig. 3.3(a–e)]. For different blocks, its filters range from convolutions with a stride of (1,1) or (2,2), convolutions with a dilation rate of (2,2), to transpose convolutions with a stride of (2,2). The activation functions after each layer is Rectified Linear Unit (ReLU). And Batch Normalization (BN) is adopted to standardize the inputs for the next layer [119]. The skip connection

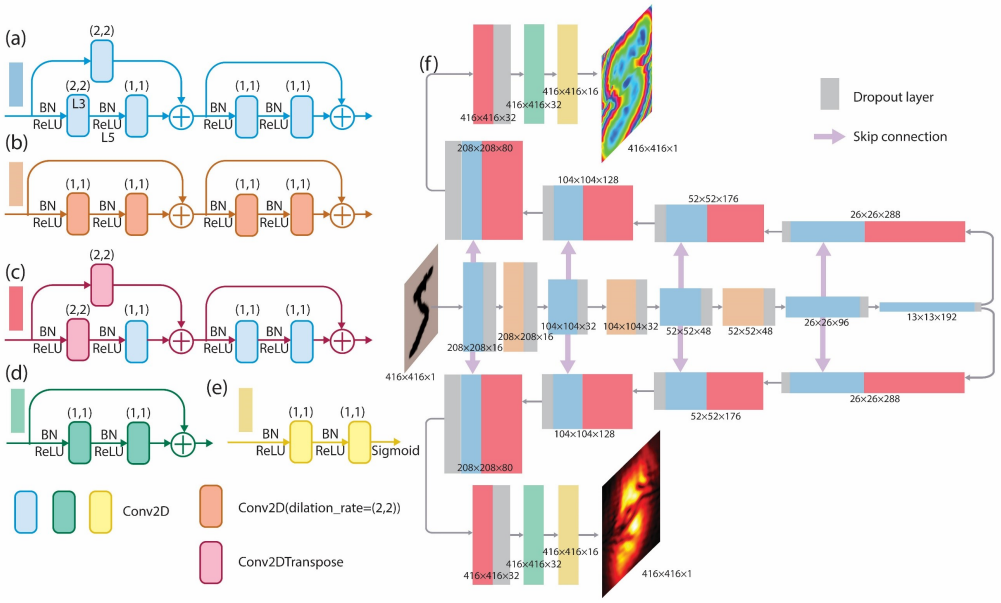

Figure 3.3: The architecture of the NO model consists of (a) down sample, (b) dilation, (c) up sample, (d) residual and (e) convolution blocks. Different blocks are represented in a single rectangle of its corresponding color. (f) The detailed structure of the NO model is built with all the above blocks with dropout layers denoted by a gray block and skip connections represented by a purple arrow. The numbers on the top or bottom of each block such as $a \times b \times c$ indicates the size of the feature map is $a \times b$ and the feature channel is c. In each block, the convolution filter types are specified by different color in the legend at the bottom left corner, including Conv2D, Conv2D with a dilation rate of (2,2), and Conv2DTranspose. L3 and L5 means layer 3 and layer 5 for the first block, whose maximum filter activation is shown in Fig. 3.13.

inside each block is an Add function. The overall architecture of our NO model is illustrated in Fig. 3.3(f), within which the skip connection is concatenation and the dropout rate is set at 0.03 in the Dropout Layer. The kernel size of all the layers is (3, 3). The optimizer is Adam [120] with a learning rate of 0.001.

The input layer is the desired observation (holographic image) with a size of 416 × 416. Then it passes through down sample and dilation blocks. The channel size is specified on each block [Fig. 3.3(f)]. Within each down sample block, the size of the image is reduced by a factor of 2 along each dimension. The size reduces to (13, 13) at the final down sample block with a channel size of 192. Next, the block separates into two branches. The phase branch [top path in Fig. 3.3(f)] contains five up sample blocks followed by a residual and a convolution block. The anterior four up sample blocks are concatenated with skip connection from the down sample path so as to solve the diminishing gradient issue and extract local and global features [121]. The amplitude branch, [lower path in Fig. 3.3(f)] consists of the same structure as phase branch.

The objective function to be minimized during training is the mean absolute error (MAE),

$$Loss = \sum \frac{1}{wd} ||P - T||, \tag{3.5}$$

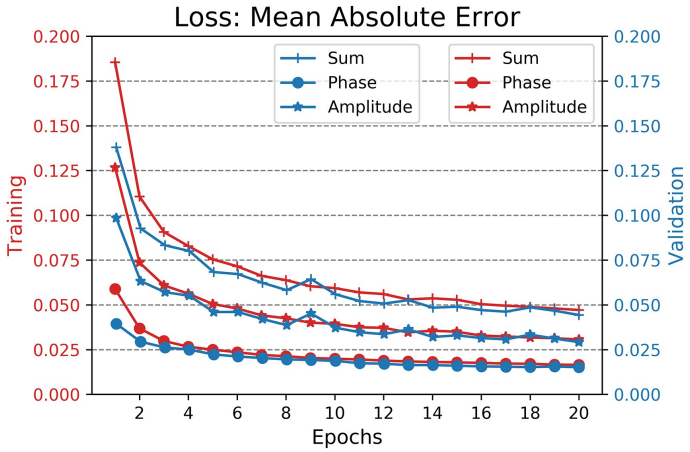

Figure 3.4: Training history over 20 epochs. Red line denotes training loss while blue line represents validation loss with respect to epochs.

where w and d denote the width and height of the output, P means the prediction output, and T represents the true output label, which is normalized in the range of [0,1]. The final MAE sums over the MAE's from two branches with a ratio of 1:1. The MAE is actually the pixel-wise average deviation for the prediction field. Therefore, the MAE is a raw value. In order to represent the percentage of deviation, we use a term, mean absolute deviation (MAD), i.e., MAD = MAE/span. The denominator is chosen as 1.0 given both amplitude and phase are normalized to a span of 1. Here, MAE is an absolute value whereas MAD is a percentage ratio.

We selected 8000 images from the Caltech101 dataset and resized them into 416×416 pixels. For each resized image H, we used Equation 3.3 to calculate the corresponding metasurface pattern M as its label, which led to 8000 pairs of datasets. The focusing distance from metasurface M to holographic image H was 750 μm. Each unit pixel had a size of 500 $\times$ 500 nm^2. Hence, both the target holographic image and metasurface had a size of 213 $\times$ 213 μm^2. The dataset was split into training and test dataset with a 9:1 ratio, which means 7200 pairs for training and 800 pairs for testing. Within the training data, we also took 10% of them as validation. Therefore, in every training epoch, 6480 pairs were used for training while 720 pairs were set aside for validation.

We trained the NO for 20 epochs with our dataset. The training history [Fig. 3.4] suggested that the loss had already reached a plateau from epoch 8 and onwards. At epoch 20, the training MAE reached 0.0471 (MAD of 4.71%) and the validation MAE was 0.0443 (MAD of 4.43%), implying that the NO was not overfitting. The test MAE was 0.0457 (MAD of 4.57%), suggesting a good bias and variance. Since Dropout layer will be deactivated in the evaluation process, the validation and test loss were slightly lower than the training loss.

In order to validate the novelty and impact of the NO, I compared it with three other

methods, namely, gradient-based optimization, global genetic algorithm and coupled-mode theory. The implementation of these three methods are detailed in Ch. 3.5 . I took the "HKU" letters as the benchmark testing sample with the visual comparison displayed in Ch. 3.5.4. The comparison results were listed in Table 3.1, which shows a greater edge of our NO over other methods.

Table 3.1: Comparison table of four different methods to design a metasurface. All these methods were tested on the same CPU of AMD EPYC 7302P 3.0GHz (16C/32T, 128M CACHE). The training time is not applicable to gradient-based optimization, global genetic algorithm and coupled-mode theory. 1 iteration time denotes the time to run one forward simulation and one backward propagation to update the parameters. Running time records total forward propagation time to deliver a result. For example, the trained NO only requires one forward propagation to deliver the prediction, while the other three methods need multiple iterations to output the results. Error denotes the deviation percentage between simulated intensity and desired intensity averaged across pixels. Metasurface and holographic image size denotes the number of units we adopted.

Method	Training time	1 iteration time	Running time	Error	Metasurface size	Holographic image size
Neural optimizer (NO)	30 hrs	N/A	1 s	1.38%	416×416	416×416
Gradient-based optimization	N/A	1146 s	1.6 hr	10.69%	416×416	100×100
Global genetic algorithm	N/A	1013 s	5.6 hr	19.90%	416×416	100×100
Coupled-mode theory	N/A	106 s	5.9 hr	70.72%	52×52	25×25

After nearly 30-hour training, our NO could deliver a metasurface design in about 1 s with a 1.38% error in our benchmark testing of "HKU" letters. In comparison, other three methods had to start from the beginning for each new input with an increased error. When the number of metasurface reaches 20, the total running time of gradient-based optimization is already much longer than our NO. Besides, in the three methods other than NO, I had to reduce the metasurface size or holographic image size due to the memory limit of our workstation. I acknowledged that the three methods other than NO may require hyperparameter tuning to increase efficiency and reduce error. But it was beyond the scope of this manuscript. And the benchmark testing dataset should contain a bundle of testing images rather than a single image of "HKU" letters. However, it would be too heavy a computation burden to prepare a large dataset for the community to duplicate while only one image would be easier to work with. I also validated our NO using a different open source dataset detailed in Ch. 3.6.

In conclusion, our NO has exceeded other three methods in terms of running time, error rate, metasurface and holographic image sizes.

3.5 Benchmark comparison

In Table 3.1, I compared our NO against three other inverse design methods: gradient-based optimization, global genetic algorithm and coupled-mode theory. The detailed implementation of these three methods is listed as follows.

3.5.1 Gradient-based optimization

For gradient-based optimization, I first initialized the complex map of the CGH. Then I apply the Huygens-Fresnel equation to compute the far-field intensity. Next, the loss function is obtained from the deviation of computed and desired intensity distribution. After that, the gradient descent algorithm is applied from the loss function to each pixel on the CGH. Finally, I output the optimal solution after multiple loops.

The tunable parameters are the complex modulation of nano-antennas. Given our CGH has 416×416 nano-antennas, the parameters are of size $2 \times 416 \times 416$, where 2 represents the amplitude and phase modulation.

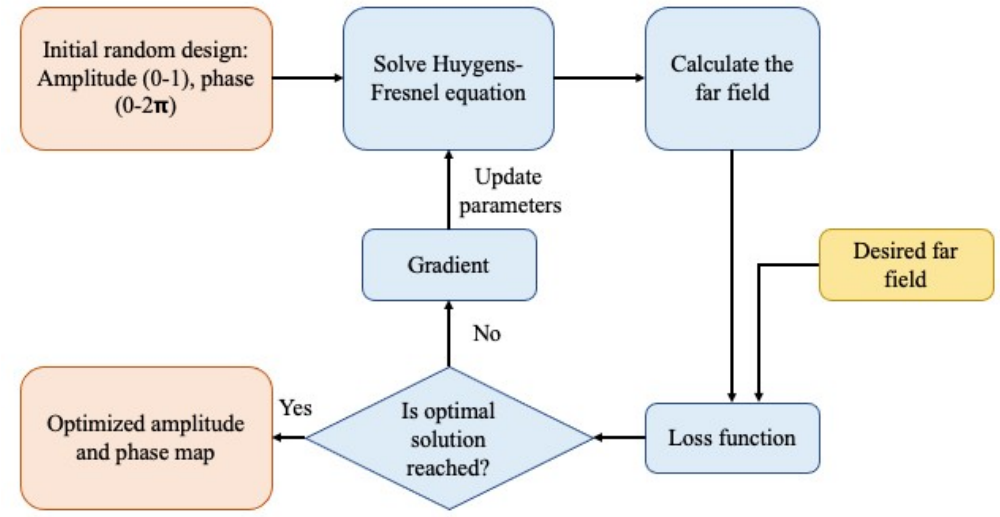

Figure 3.5: Metasurface design procedure using gradient-based optimization. Firstly I initialize a random complex map. Then the far-field is computed. Next, I get the loss function by comparing the computed far-field with the desired far-field. Then the gradient descent step is applied to each parameter. At last, I output the converged solution.

We initialize the amplitude and phase as A and P by a 416×416 matrix. The near field of the CGH could be written as $E = A \cdot exp(1j \cdot P)$. The focal length is 750 μm. The holographic image intensity distribution is calculated through the Green function, which is

$$H(i,j) = \sum_{i,j} \frac{exp(ik \cdot R_{ij}(x,y)}{R_{ij}(x,y))} \cdot M(x,y), \tag{3.6}$$

where M(x,y) denotes the complex modulation at position (x,y) on a metasurface M, k represents the wavenumber, H(i,j) is the intensity information at position (i,j) for a given holographic image H, and $R_{ij}(x,y)$ is the pixel-wise distance between M and H, which in mathematical form is $|\overrightarrow{OM}(x,y) - \overrightarrow{OH}(i,j)|$ with O being the coordinate origin.

The loss function is defined as

$$Loss = \sum ||D - H||, \tag{3.7}$$

where D is the desired intensity, H is the computed intensity.

Each parameter follows the gradient descent rule, which is

$$p := p - \frac{dL}{dp}, \tag{3.8}$$

where p denotes the amplitude or phase of one unit-antenna.

3.5.2 Global genetic algorithm

The global genetic algorithm we adopted here is similar to this reference [122]. I firstly initialized 5 different complex maps of the CGH as the parent generation. Then I randomly tuned each pixel's amplitude and phase modulation with a descending mutation rate. Next, each parent will deliver 5 children, and the children with the highest fitness score will survive. The new 5 children will form the new parent generation. Finally, I output the optimal solution after multiple loops.

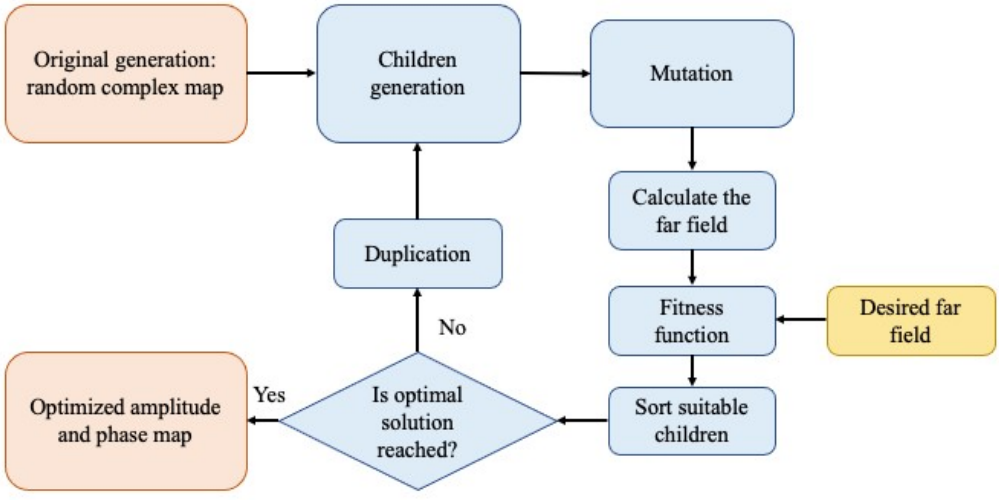

Figure 3.6: The flowchart of the genetic algorithm. It starts with the random complex map as the original generation. Then I apply the mutation to the current generation. Next, I sort the children generation according to their fitness scores. If the solution is not converged, I duplicate the children generation and start the loop again. At last, I output the best solution.

In initialization, the degree of freedom is 2N when the metasurface has N elements because two numbers are required to represent the complex modulation. The mutation rate is set at 0.3 for the first generation and drops down to 0.05 thereafter. The definition of mutation rate here is the possibility of one element changing its amplitude or phase. Then, I calculated the fitness score of each child. The desired intensity is denoted as D. The computed intensity H follows the same calculation procedure in previous methods. The fitness score is defined as

$$F = -\sum ||D - H||. \tag{3.9}$$

If the difference between the desired intensity D and computed intensity H is smaller, the child in question will perform better, which mathematically means a higher fitness score. Each child will generate 5 new children in the mutation process. The total 6 individuals will be evaluated against the fitness score with the top one being the new parent.

3.5.3 Coupled-mode theory

I also adopted coupled-mode theory (CMT) to inversely design the plasmonic metasurface. The method is quite similar to this reference [123]. Here I only considered the far field response of each Au antenna. The LCP to RCP is neglected to ease computation.

Firstly, I randomly initialized a metasurface with M elements. Then I solved the coupled-mode theory to compute the far-field intensity distribution. Next, the loss function was obtained, and gradient descent was applied to update each metasurface element. At last, I output the optimized solution after multiple loops.

Then, I ran an FDTD simulation for an Au antenna with a width of 150 nm and a varying length from 50 nm to 250 nm. Then the resonance frequency and quality factor were recorded to build a one-to-one mapping from configuration to resonance frequency. The tunable parameters $\vec{p}$ in the CMT model is set as $p = (\omega_1, \omega_2, ..., \omega_M)$, where M denotes the number of Au antennas.

Next, the CMT method is applied to calculate far-field radiation based on the initial parameters. The flow is detailed in literature [123], and I also briefed it as follows. I built the $M \times M$ coupling matrix as Ω. The diagonal elements are the resonant frequencies. The off-diagonal elements is calculated as

$$\Omega_{mm'} = \frac{1}{2}\sqrt{\frac{\omega_m \omega_{m'}}{Q_m Q_{m'}}} Y_0(k_0 |\vec{r_m} - \vec{r_{m'}}|), \tag{3.10}$$

where Y_0 is the Bessel function of the second kind, k_0 being the wavenumber.

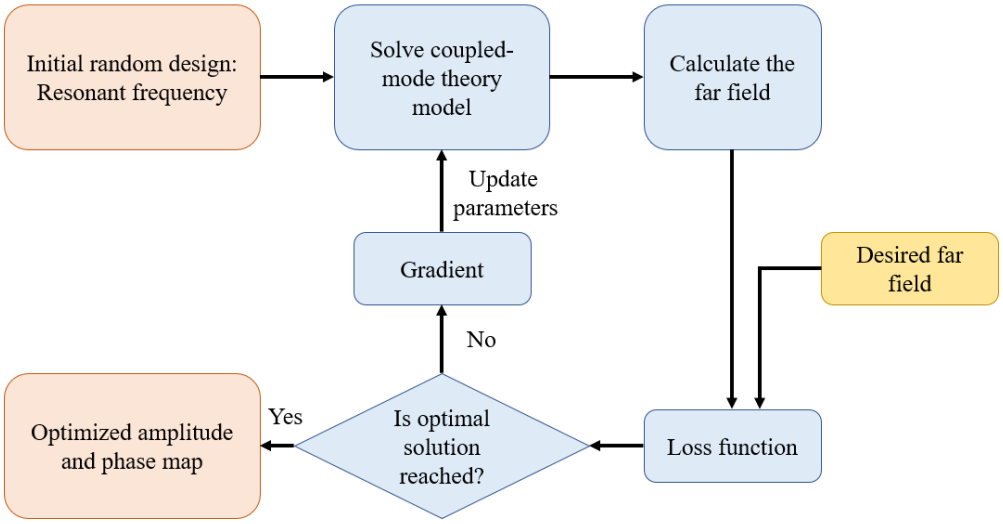

Figure 3.7: Metasurface design procedure using coupled-mode theory optimization. Firstly I initialize a random resonant frequency with each frequency mapping to a specific nano unit configuration. Then I compute the far-field to compare with the desired far-field to get the loss function. Next, the gradient descent step is applied to each parameter. At last, I output the converged solution.

The incoming and outgoing waves are represented by N-by-1 vectors S_{in} and S_{out}. The far-field intensity at location $r' = (x', y', z')$ is

$$F(\omega_0, \vec{r'}) = \sum |W(\vec{r'}) S_{out}(\omega_0)|^2, \tag{3.11}$$

where $W(r) = (e^{-jk_1 \cdot r'}, e^{-jk_2 \cdot r'}, ..., e^{-jk_N \cdot r'})$ is the propagation phases of each channel.

The loss function and the updating rule are the same as the gradient-based optimization in Ch. 3.5.1.

3.5.4 Visualization comparison of far-field radiation

For all the four methods listed in Table 3.1 in this thesis, I have presented their visual comparison in Fig. 3.8. The input image was the "HKU" letters but of different resolution sizes due to the computation limit. Then the metasurface was designed through the four methods. At last, we computed the holographic image of each optimized metasurface and plot them in the second row in Figure S1. Our NO has the clearest illumination and highest resolution size under the same computational source.

3.6 Experimental verification

The metasurface consisted of Au blocks coated on a glass substrate. The Au blocks were fixed with a height h of 80 nm and a width w_x of 200 nm. The length of the Au block ranged from 200 nm to 400 nm, and the orientation angle tuned from 0 to π to achieve

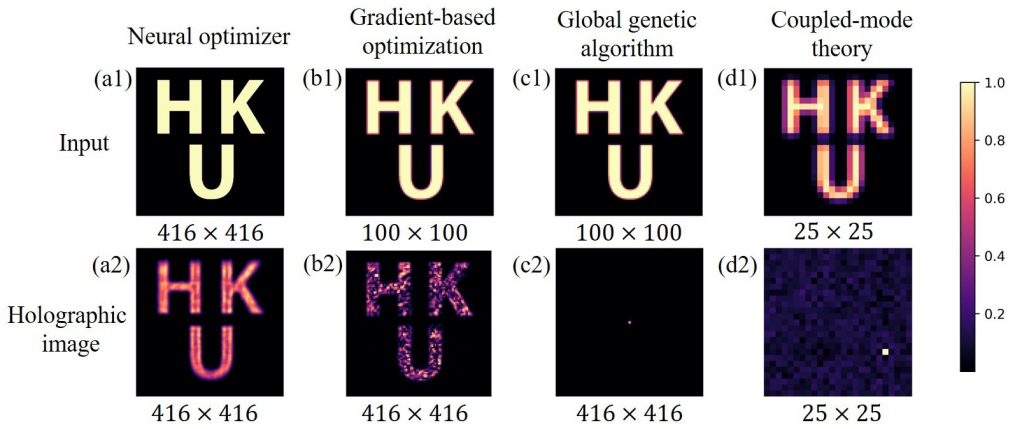

Figure 3.8: Visualization comparison of the four methods from Table 3.1, which are (a1, a2) NO, (b1, b2) gradient-based optimization, (c1, c2) global genetic algorithm, and (d1, d2) coupled-mode theory. The first row presents the input images for different methods. Their resolution sizes are captioned below. The second row shows the holographic images of the optimized metasurface with their resolution size captioned below.

the 2π phase modulation according to Pancharatnam–Berry phase [79]. The design of the Au block to achieve complex modulation was well presented previously [88]. In brief, the Au block illuminated by a LCP light functioned as a birefringent crystal, and part of the LCP light changed into RCP light [inset in Fig. 3.9]. The output can be formulated as

$$S_{out} = \langle R|\Gamma(-\alpha)\hat{W}\Gamma(\alpha)|L\rangle, \tag{3.12}$$

where R and L respectively denote RCP and LCP; $\Gamma(\alpha)$ represents the rotation matrix; α indicates the orientation angle of the Au block; and $\hat{W}$ is the transformation matrix.

The experimental setup is shown in Fig. 3.9, which is similar to Fig. 2.5. The PBS is to convert the incoming beam into a polarized beam. The LCP is to convert a polarized beam into an LCP beam, while the RCP is to filter out the RCP component of the transmitted beam. The lasers we used are at three separate wavelengths, namely, 808 nm, 1030 nm, and 1200 nm. The CCD is from IDS imaging with an item number of UI-1240SE-M-GL. LCP is a polymer circular polarizer from Thorlabs (CP1L1064). RCP is also a polymer circular polarizer from Thorlabs (CP1R1064). The selection criteria is that they are readily available from our lab or Thorlabs.

The electric field of the filtered RCP component is

$$E_{out} = i \cdot sin[\frac{kh(n_o - n_e)}{2}] \cdot exp(i[\frac{kh(n_o + n_e)}{2} + 2\alpha]), \tag{3.13}$$

where k is the wavevector, n_o and n_e are the ordinary and extraordinary index of the Au block respectively. I have run a full-wave simulation for the Au antenna through Lumerical FDTD. The length of the antenna was specially picked to achieve an amplitude range of 0 to 1, while the orientation was tailored to exert a full range 2π modulation.

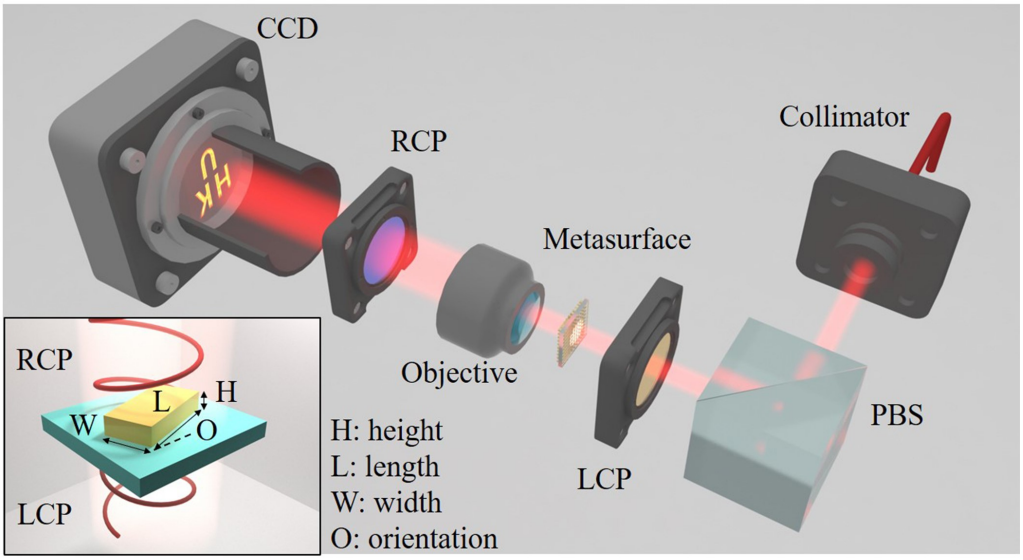

Figure 3.9: Experimental setup. PBS: polarization beam splitter. LCP/RCP: left/right circularly polarized light filter, containing a $\lambda/4$ waveplate and a polarizer. CCD: charge-coupled device camera. Inset: unit structure of the metasurface. It consists of Au block on a glass substrate.

In order to create the targeted observation pattern, the Au antenna with different lengths and orientations would be arranged to form a specific pattern. Accordingly, we adopted the NO to efficiently design such pattern. The metasurface was then fabricated through electron-beam lithography (EBL). I built an experimental platform to test the fabricated metasurface [Fig. 3.9]. The 1030-nm incident light was prepared with an LCP polarization, and then illuminating on the metasurface. Next, the modulated light passed through the RCP filter, finally cast a desired holographic image on the camera.

I selected two images to test the concept. None of them has been used to train the NO. They were "HKU" letters [Figure 4(a)], Chinese character "guang" (meaning 'light') [Fig. 3.10(b)]. The theoretical complex field of the metasurfaces [row I in Fig. 3.10] and the corresponding holographic image [row II in Fig. 3.10] were calculated through Equation 3.3. It took 450 s to compute through MATLAB on a CPU of Intel i5 8500 4.1GHz. In contrast, with the same CPU, the NO prediction only took about 1 s. Before training, the prediction of the neural network was purely noise without meaningful information [column a2, b2 in Fig. 3.10]. After training, the generated holographic image matched well with the theoretical one [column a3, b3 in Fig. 3.10]. Then we fabricated the metasurface featuring with the complex modulation for testing . The experimental holographic image matched well with our theory [Fig. 3.10].

Due to the ohmic loss and amplitude modulation, the transmitted power, in the case of "HKU" letters, was only 1.87% of the incoming power. However, the experimental results sufficed to demonstrate the concept of NO. And we will consider to turn to dielectric metasurface in the future. The issue of the blurred letter "U" comes from

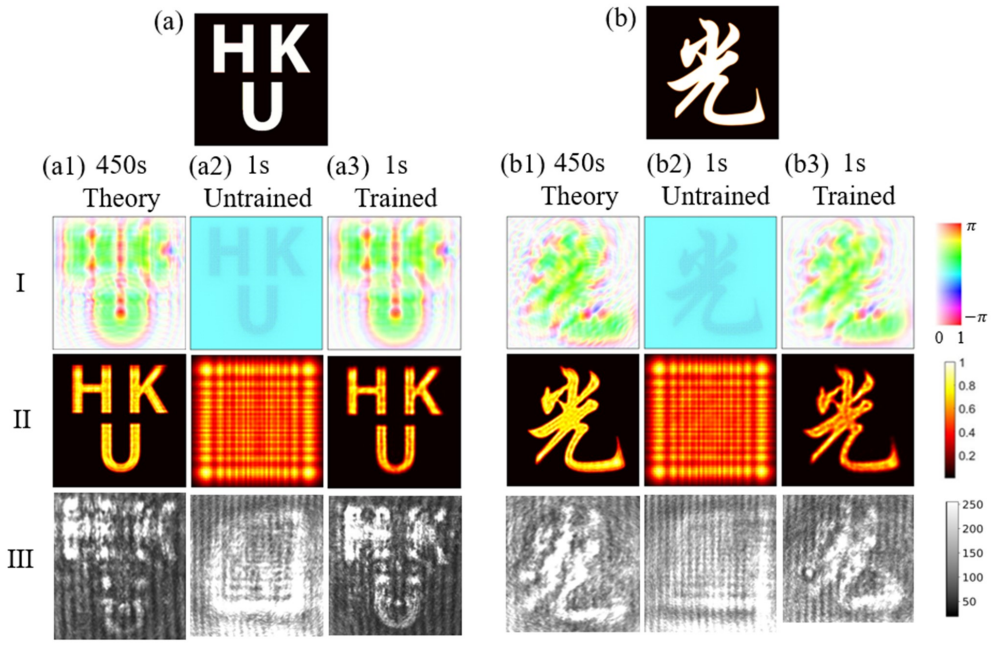

Figure 3.10: Experimental results of NO. Row I: complex field of metasurface's pattern. Row II: simulated holographic image with a distance of 750 µm from the metasurface. Row III: holographic image in experiments. The observed holographic images are (a) "HKU" letters, and (b) "guang" character. (Column a1, b1) Theoretical computation as in Equation 3.3 with a computation time of 450 s. (Column a2, b2) Prediction through the untrained NO with 1 s. (Column a3, b3) Prediction through the trained NO with 1 s. The training process takes 30 hours to complete.

the fabrication imperfection due to the unstable performance of the EBL machine. The optical components we selected also affected the results. For example, the LCP and RCP is not of 100% efficiency. The CCD will also suffer from background noise. These optical components can be refined to futher improve our results.

The SEM picture of the metasurfaces used in Fig. 3.10(a) is shown in Fig. 3.11. Fig. 3.11(a) illustrates the metasurface which is theoretically computed to deliver the holographic image of "HKU" in Fig. 3.10(a1). Fig. 3.11(b) shows the metasurface which is predicted by an untrained neural optimizer to deliver the holographic image of in Fig. 3.10(a2). Fig. 3.11(c) shows the metasurface which is predicted by the trained neural optimizer to deliver the holographic image of "HKU" in Fig. 3.10(a3).

The SEM picture of the metasurfaces used in Fig. 3.10(b) is shown in Fig. 3.12. Fig. 3.12(a) illustrates the metasurface which is theoretically computed to deliver the holographic image of "guang" in Fig. 3.10(b1). Fig. 3.12(b) shows the metasurface which is predicted by an untrained neural optimizer to deliver the holographic image of in Fig. 3.10(b2). Fig. 3.12(c) shows the metasurface which is predicted by the trained neural optimizer to deliver the holographic image of "guang" in Fig. 3.10(b3).

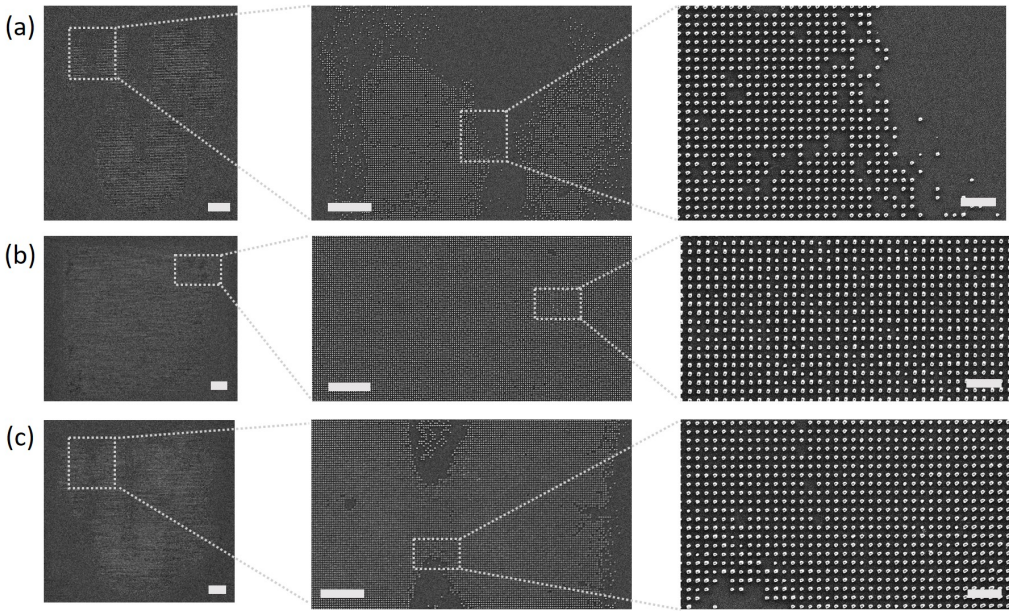

Figure 3.11: The SEM picture of the metasurfaces used in Fig. 3.10(a). (a) Metasurface that is theoretically computed in 3.10(a1). (b) Metasurface that is predicted through an untrained neural network in 3.10(a2). (c) Metasurface that is predicted through the trained neural optimizer in 3.10(a3). Scale bar: 20 μm in column I, 10 μm in column II, 2 μm in column III.

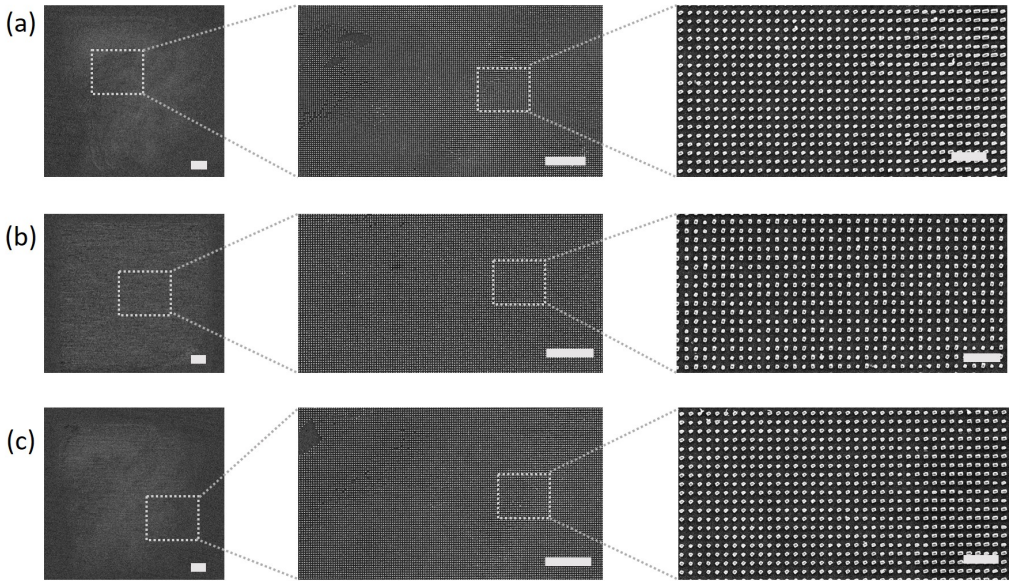

Figure 3.12: The SEM picture of the metasurfaces used in Fig. 3.10(b). (a) Metasurface that is theoretically computed in 3.10(b1). (b) Metasurface that is predicted through an untrained neural network in 3.10(b2). (c) Metasurface that is predicted through the trained neural optimizer in 3.10(b3). Scale bar: 20 μm in column I, 10 μm in column II, 2 μm in column III.

The training process took roughly 30 hours using Tensorflow on a CPU of AMD EPYC 7302P 3.0GHz (16C/32T, 128M CACHE).

Apart from the Caltech101 database, 2000 images from Labeled Face in the Wild (LFW) were selected as a different test pool for cross check. The test MAE of LFW database for the NO trained under Caltech101 was 0.0484 (MAD: 4.84%) on average, which was considerably good. The test images from a different database could verify that our NO generalized well across images in different datasets.

3.7 Neural network explanation

We further visualized the maximum filter activation of what has been learned by the neural network [Fig. 3.13]. The input size was 32×32. Starting from a randomly-initialized input, we used a gradient ascent algorithm to maximize the filter response [124]. The chosen filter was maximally responsive to the finalized input [Fig. 3.13]. For a shallower layer [layer 3 in Fig. 3.13(a)], it could capture simple features, including stripes or dots. When it came to a deeper layer, for example layer 5 in Fig. 3.13(b), complex features like clouds or textures were learned.

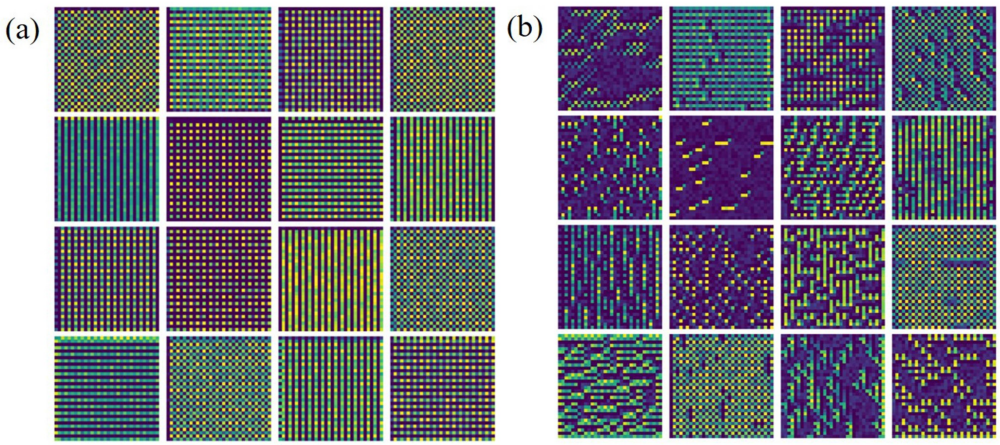

Figure 3.13: Maximum filter activation visualization. The patterns show the inputs that would maximize the 16 filter responses in (a) layer 3, the first convolutional layer in the first down sample block, and (b) layer 5, the second ReLU layer in the first down sample block. The two layers are marked in Figure 3.3. The raw image is downsized by 2 in each dimension for both layers.

To better reveal the neural network from inside, we visualized the intermediate activation output in Fig. 3.14 for the input holographic image, a shape of omega from the lab logo of Kenneth K.Y. Wong's group at HKU. The channels are the random three channels of the last convolutional layer of each specific layer. The block Do-1&3&5, Aup-1&3 (Pup-1&3), and Ares1 (Pres1) represents down sample, up sample and residual blocks (for amplitude and phase branches), respectively.

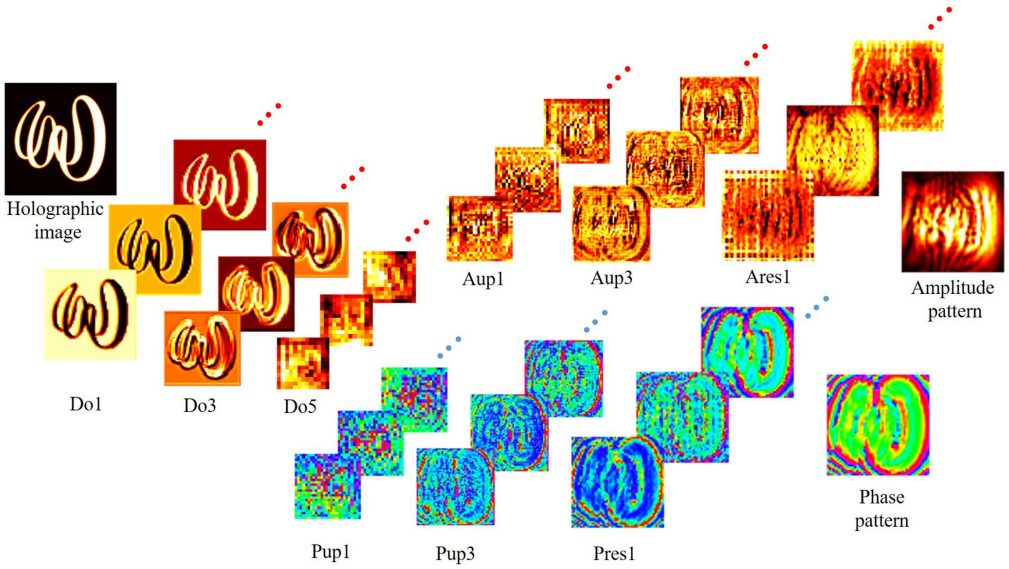

Figure 3.14: Intermediate activation map of NO when the logo of omega is inputted. Total number of channels is defined in the architecture in Figure 3.3. Only three random channels are displayed here. The chosen layer is the last convolution layer of a specific block. "Do" denotes down sample block. "Aup/Pup" represents the up sample block for amplitude and phase branch, while "Ares/Pres" means the residual block.

When a holographic image was fed into the neural network, the down sample blocks firstly extracted simple features (e.g., vertical/horizontal edges, contrast, etc.), then identified complex features (e.g., thickness of a stripe, etc.). After that, the up sample branches utilized such information to reconstruct the amplitude and phase pattern, which will lead to a desired holographic image.

3.8 Discussion

Here, a neural optimizer is demonstrated to design a metasurface within 1 second to project a holographic image in a lensless modality. This neural network proves to have a good generalization with an error rate below 5% across different datasets.

Machine learning has been used a lot to design a metasurface in literature. However, our research is unique in that it explores the usage of neural network in a specific lensless modality, which is promising in near-eye display applications. It accelerates the pattern design process to allow for a real-time display scenario.

Meanwhile, there is still a lot of room for improvement. For example, we are using a metasurface with gold nano-structure. For a display application, it requires a structure with at least 24 frame per second speed, e.g. SLM, liquid crystal. It is still a challenge to deliver the same results with SLM as the static gold metasurface does. We can still explore the neural network compression to reduce the computation time even further.

3.9 Summary

All in all, we built a NO to intelligently design a CGH for an optical lensless 2D image projection. The CGH then was fabricated into a complex-modulated metasurface, which can generate a desired holographic image without bulky lens system, much more convenient than phase-only CGH, which needs a Fourier-transform lens. Though the experimental setup here contained other optical components, including LCP, RCP, and PBS, they can be replaced by a coating to achieve a handy modality. The NO tends to generalize the optical backpropagation process in free space from an object field to the CGH. The NO was trained under the Caltech101 database and had a test loss of 4.84% for LFW database. The theoretical training results were well matched with experiments. Our NO can accelerate the inverse design process of the metasurface's pattern, from 450 s in a theoretically computing approach to 1 s enhanced by the prediction capacity. Although the time difference is negligible compared to the EBL process of nearly 10 hours, the NO can readily be applied to a complex-modulated SLM to achieve a real-time design, which we will explore further. As for time complexity, it is reduced from $O(n^4)$ for theoretical computation to $O(n^2)$ for NO. When the number of metasurface reaches above 20, the total running time of NO is significantly reduced compared to other methods, e.g. gradient-based optimization, global genetic algorithm and coupled-mode theory. Besides, the error can be lowered from more than 10% to only 1.38% for the selected testing image. Given that the complex-modulated metasurface has more sophisticated functions [44], the proposed NO may assist in lensless image projection and real-time metasurface pattern design. This work demonstrates an inverse design technique for metasurface patterning. It utilizes a neural network for an accelerated prediction of metasurface pattern.

Chapter 4

Biomedical "vision": metasurface for optical imaging

Last chapter showcased a neural network for an accelerated inverse design for metasurface patterning. However, a neural network usually requires a large amount of data to train. The training dataset, if solved by FDTD, is still a huge computation burden.

Therefore, this chapter presents an adjoint optimization method to directly optimize a structure to achieve a desired far-field radiation without any prior training data. The adjoint method is applied to tailor a fiber facet to build an endoscope which can achieve an extended depth of focus for vessel imaging through optical coherence tomography. Current issues in endoscopic optical imaging include limited penetration depth, optical aberration, trade-off between lateral and transverse resolution, etc. Here I designed subwavelength structures on the beveled fiber facet, which will be etched by focused ion beam (FIB). The design process was accomplished through adjoint optimization. Firstly, a geometrical optics solution has been proposed for benchmark comparison. The structures vary in geometry to reflect the incident light with a phase modulation ranging from 0 to 2π. Then, a wave optics solution will be calculated to obtain the grating structure on a fiber facet, which will produce an electric field response with an extended depth of focus. By applying these two pattern, the reflected beam will be tested of its beam profile. If it can achieve a customized focus depth surrounding the fiber, I will integrate the meta-fiber into an imaging modality. When the beam is transmitted out of the probe and scattered by the tissue and propagate back along the fiber, I will collect the signal and conduct the Fourier analysis to reconstruct the tissue. The tailored fiber endoscope will promote biomedical imaging and disease diagnosis for vessels regarding clinical needs.

4.1 Introduction

Optical imaging provides a non-invasive technique to acquire a detailed image of the tissue through photons across a wide electromagnetic spectrum. Its modalities include

optical coherence tomography (OCT) [42], photoacoustic (PA) imaging [125], Raman spectroscopy [126], super resolution microscopy [127], etc.

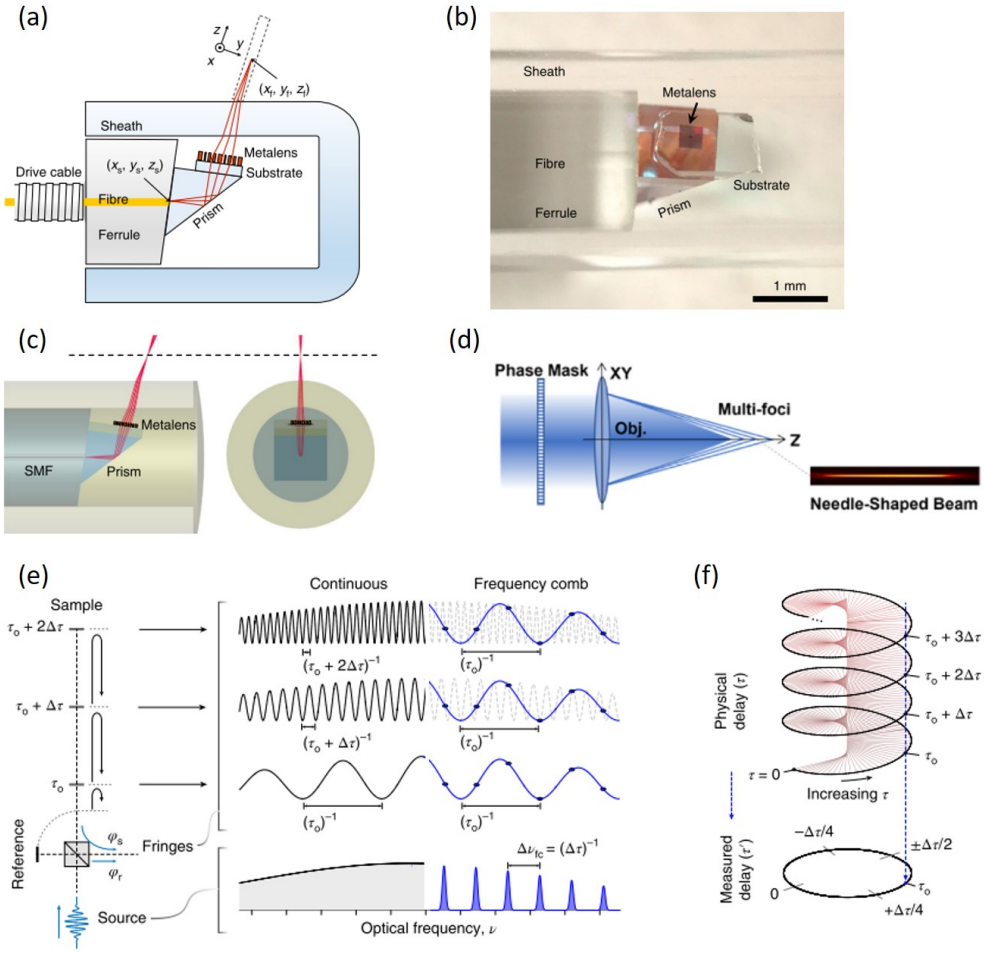

Figure 4.1: Recent developments in OCT. (a, b, c) A metasurface-integrated endoscope for an enhanced imaging depth [42]. (a) Schematic figure of the endoscope probe. (b) A photo of the fabricated device. (c) Focusing property of the OCT. (d) The phase mask to generate a needle-shaped beam [128]. (e, f) High speed OCT with circular interferometric ranging technique [129]. (e) Base delay can be confirmed from the same spectral modulation frequency. (f) Circular mapping from the measured delay to the physical delay.

Huge effort has been spent to improve the above imaging modalities in terms of speed [129], resolution [130], miniaturization [131], etc. In OCT field, a metasurface has been designed to compenstate the phase shift caused by the endoscope sheath [Fig. 4.1(a,b,c)], thus achieving a diffraction-limited focus with a reduced aberration [42]. Another group proposed a diffraction optical element (DOE) [Fig. 4.1(d)], when fabricated according to a specific phase pattern and positioned in front of an objective in an OCT modality, could form a needle-shaped beam with a 50-fold Rayleigh range [128]. Such a long depth of focus (DOF) can improve the imaging depth and resolution. A circular coherent ranging architecture is demonstrated to compress sparse physical voxels (e.g.

air) into one measurement set [Fig. 4.1(e,f)]. Compared with Fourier-domain OCT, this method can reach an imaging speed of 7.5 volumes per second through the sparsity compression [129], which will play an important role in dynamic biomedical imaging.

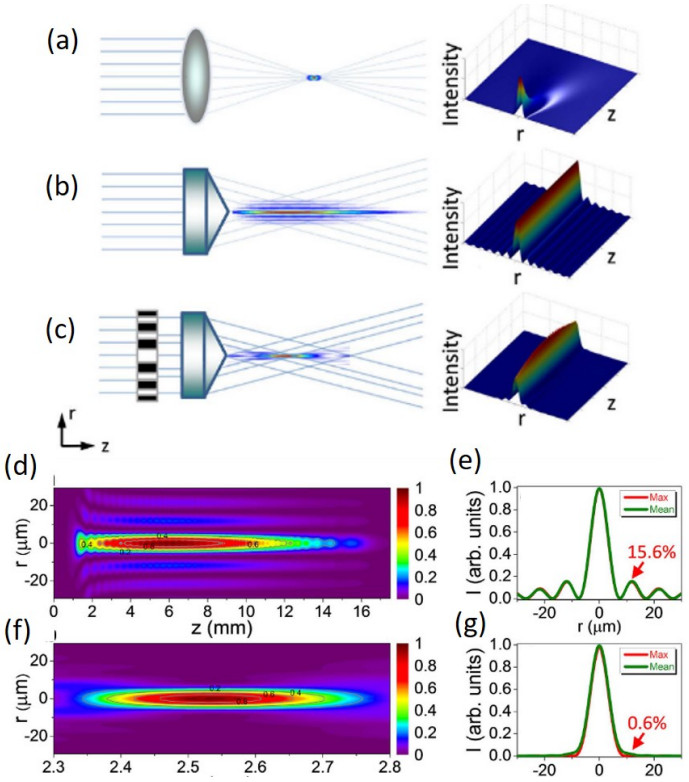

Figure 4.2: Comparison between Bessel beam and EDOF [132]. Illustration of intensity profile for (a) Gaussian beam, (b) Bessel beam, and (c) tailored EDOF. Simulated intensity profile of (d, e)Bessel beam and (f, g) tailored EDOF.

Among the all the factors affecting imaging quality, extended depth of focus (EDOF) demonstrates great importance in improving imaging depth and resolution [133]. In this regard, Bessel beam and Airy beam has been frequently adopted as a penetration beam due to their diffraction-free features [134]. However, such features come with high-energy side lobes, which will cause background noise and photobleaching [134]. Therfore, the community proposes various methods to tackle this problem. The comparison between Bessel beam and EDOF is displayed in Fig. 4.2. One group designed a phase mask on a metasurface to achieve an EDOF across visible spectrum with a suppressed achromatic aberration [135]. Another group proposed a DOE, whose phase pattern allocates the incident beam into multiple focal points along the axial axis, to generate an EDOF with diffraction-limited beam width [128].

However, the above designing methods, e.g. metasurface geometry sweeping, phase mask designing, are actually a forward process, which may pose high computation burden [134] but lead to a decreased focusing efficiency [135]. Recently, some inverse design methods have been applied in metasurface designing [136]. This group

applied adjoint optimization to design a metalens with a 24% efficiency higher in holography focusing than conventional periodic unit-cells [136]. Another group also utilized adjoint method to inverselt design a cylindrical meta-grating with an EDOF at 625 nm wavelength [133].

While EDOF plays an important role in imaging quality, the community also looks for miniaturized device for a greater versatility. One group reported a graded-index (GRIN) fiber, when connected with a single mode fiber (SMF), can provide an ultra-small probe for a better lateral resolution [137]. This article claimed nano technology has enabled imaging probes with an outstanding performance within a limited size [138], e.g. magnetic particles for a higher sensitivity in magnetic resonance imaging (MRI), quantum dots as a robust fluorescent tags due to their optical and chemical stability, high atomic number elements for an enhanced computed tomography (CT), etc.

All in all, while numerous research is going on about probe miniaturization and nano-photonic device for EDOF, the research for OCT with an SMF probe, whose facet is inversely designed through adjoint optimization to achieve an EDOF, is still under-explored. In this chapter, I will demonstrate the usage of adjoint optimization to design the pattern on a SMF, which can be readily adopted by OCT or PA imaging.

4.2 Theory for adjoint optimization

This section presents the working priciple of adjoint optimization.

4.2.1 Mathematical derivation

Adjoint method was well explained in literature [139][140]. It provides a way to calculate derivatives with less computation burden. Such derivative represents the sensitivity of the output results to the input parameters. Given the emerging problems nowadays usually involve a large amount of parameters, e.g. machine learning models with millions of parameters to tune, mechanical simulation with tons of voxels as input, etc., adjoint method is an excellent way out leading to an optimized parameters set within a limited time.

Firstly, we have a set of field values $\vec{x} \in R^{n_x}$, represented as $[x_1, x_2, ..., x_{n_x}]^\intercal$. And we also have a set of parameters $\vec{p} \in R^{n_p}$, represented as $[p_1, p_2, ..., p_{n_p}]^\intercal$. In order to calculate the field value, we usually have to run a simulation or a forward computation, which could be written as

$$f(\vec{x}, \vec{p}) = 0, \tag{4.1}$$

whose dimension is $\mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_x}$. For a machine learning problem, f is simply a forward calculation. For an EM problem, f usually conveys an FDTD simulation with a certain structure, mathematically written as $\vec{p}$, to get the EM response $\vec{x}$.

Then we will define a loss function or, in a different name, a measure of merit, as

$$g(\vec{x}, \vec{p}) : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}, \tag{4.2}$$

which we would like to minimize. A machine learning problem would require the loss to be as small as possible. A mechanical or EM problem would expect some structure response to be minimized or maximized. For a maximization problem, we could add a minus $(-)$ signal in Eq. 4.2, and it will become equivalent to a minimization problem.

To minimize the loss in Eq. 4.2, gradient descent algorithm [141] is commonly applied to update the input parameters. That's where the magic of adjoint method comes into play. Since it could calculate the gradient of all parameters in the same complexity as a forward computation, it could accelerate the convergence of parameters in a way to minimize the loss function, which is equivalent to one update of a machine learning model.

Then, I gave the total derivative of the loss function as follows

$$\frac{dg}{d\vec{p}} = \frac{\partial g}{\partial \vec{p}} + \frac{\partial g}{\partial \vec{x}} \cdot \frac{d\vec{x}}{d\vec{p}}, \tag{4.3}$$

whose dimension is

$$\begin{aligned}[\frac{dg}{dp_1}, \frac{dg}{dp_2}, \ldots, \frac{dg}{dp_{n_p}}] =& [\frac{\partial g}{\partial p_1}, \frac{\partial g}{\partial p_2}, \ldots, \frac{\partial g}{\partial p_{n_p}}] \\ &+ [\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_2}, \ldots, \frac{\partial g}{\partial x_{n_x}}] \cdot \begin{bmatrix} \frac{dx_1}{dp_1} & \frac{dx_1}{dp_2} & \cdots & \frac{dx_1}{dp_{n_p}} \\ \frac{dx_2}{dp_1} & \frac{dx_2}{dp_2} & \cdots & \frac{dx_2}{dp_{n_p}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dx_{n_x}}{dp_1} & \frac{dx_{n_x}}{dp_2} & \cdots & \frac{dx_{n_x}}{dp_{n_p}} \end{bmatrix}.\end{aligned} \tag{4.4}$$

Since the loss function $g(\vec{x}, \vec{p})$ is already defined in advance in Eq. 4.2, it is not hard to get the partial derivative of $\partial g/\partial \vec{p}$ and $\partial g/\partial \vec{x}$. The hardest part is the calculation of $d\vec{x}/d\vec{p}$.

For traditional forward calculation method, we add a small deviation

$$p_i \to p_i + \delta \tag{4.5}$$

to only one parameter. Then we run the simulation to get the field value change as

$$\vec{x_{p_i}} \to \vec{x_{p_i}} + \delta \vec{x_{p_i}}. \tag{4.6}$$

Thus we have one column vector of $d\vec{x}/d\vec{p}$ as $\vec{x_{p_i}}/\delta$. Since we have n_p parameters, it would require n_p simulations to get the total derivative $d\vec{x}/d\vec{p}$. When n_p becomes increasingly large, which is normal for machine learning, mechanical or EM simulation, etc., it would be a huge computation cost for only one update.

Adjoint method can calculate the derivative $d\vec{x}/d\vec{p}$ within a computation cost no more than a forward simulation. Since a simulation $f(\vec{x},\vec{p})=0$ is satisfied everywhere, we could write the total derivative as follows

$$\frac{df}{d\vec{p}}=\frac{\partial f}{\partial \vec{x}}\cdot\frac{d\vec{x}}{d\vec{p}}+\frac{\partial f}{\partial \vec{p}}=0. \tag{4.7}$$

Let's assume $\partial f/\partial\vec{x}$ is everywhere non-singular. The field derivative to the parameters is

$$\frac{d\vec{x}}{d\vec{p}}=-(\frac{\partial f}{\partial \vec{x}})^{-1}\cdot\frac{\partial f}{\partial \vec{p}}. \tag{4.8}$$

Then we apply Eq. 4.8 to Eq. 4.3. We will have

$$\frac{dg}{d\vec{p}}=\frac{\partial g}{\partial \vec{p}}-\frac{\partial g}{\partial \vec{x}}\cdot(\frac{\partial f}{\partial \vec{x}})^{-1}\cdot\frac{\partial f}{\partial \vec{p}}. \tag{4.9}$$

In Eq. 4.9, $\partial g/\partial\vec{x}\cdot(\partial f/\partial\vec{x})^{-1}$ could be replaced as $\lambda^{\intercal}$, which is exactly the solution of

$$(\frac{\partial f}{\partial \vec{x}})^{\intercal}\lambda=(\frac{\partial g}{\partial \vec{x}})^{\intercal}. \tag{4.10}$$

Eq. 4.10 is the adjoint equation, whose computation complexity is roughly the same as a simulation in Eq. 4.1. Its solution could be readily applied to Eq. 4.9 to get the final total derivative of the loss function. In the case of millions of parameters, adjoint method in Eq. 4.10 could take weeks to converge compared with 200,000 years as in conventional forward simulation in Eq. 4.5 [142].

If the simulation $f(\vec{x},\vec{p})=0$ is a linear process, it could be written as

$$f(\vec{x},\vec{p})=A(\vec{p})\vec{x}-b(\vec{p})=0. \tag{4.11}$$

It is easy to have the derivative $\partial f/\partial\vec{x}=A(\vec{p})$. The adjoint equation for the linear process would be

$$A(\vec{p})^{\intercal}\lambda=(\frac{\partial g}{\partial \vec{x}})^{\intercal}. \tag{4.12}$$

If the simulation is a non-linear process, the derivative $\partial f/\partial\vec{x}$ may not have an explicit form. The adjoint equation is still in the form of Eq. 4.10.

4.2.2 Adjoint method and backpropagation

Backpropagation is an important algorithm to train machine learning models with millions of parameters [143]. Christopher Olah [142] once used a computational graph to demonstrate the mathematical tactic of backpropagation. Here I will also use a computational graph to present the similarity between adjoint method and backpropagation.

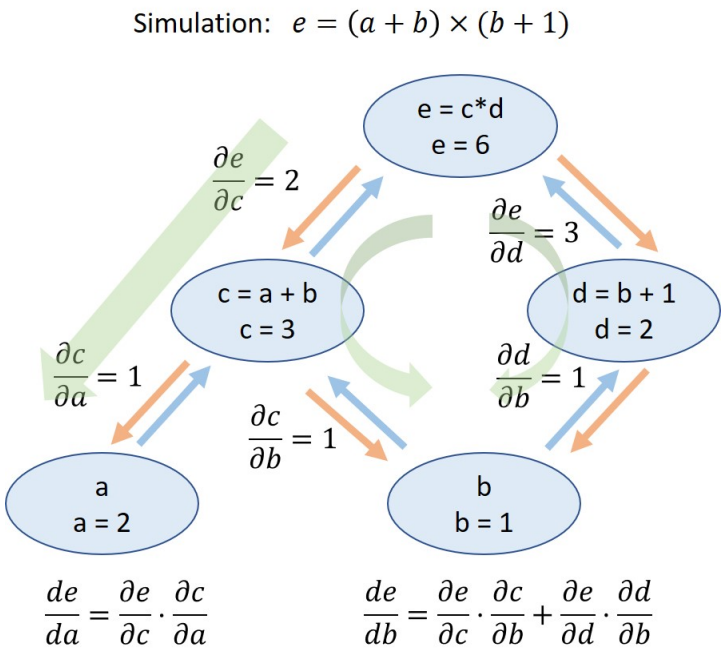

Figure 4.3: Computational graph of a simulation process $e = (a + b) \times (b + 1)$.

Let's consider a simulation process of

$$e = (a + b) \times (b + 1), \tag{4.13}$$

where apparently e is the field value, a and b are the parameters. We assign the parameters a value of $a = 2$ and $b = 1$. The computational graph of the simulation in Eq. 4.13 is plotted in Fig. 4.3, where intermediary nodes are used to demonstrates the calculating process from the input parameters to the output field value.

If we want to know how the field value will change when the input parameters has a small deviation (, which is the derivative), we have two computation methods - forward-mode differentiation and reverse-mode differentiation (backpropagation) [142].

Firstly, the partial derivative is calculated between every connected node pairs. For example in Fig. 4.3, there is an intermediary process "$c = a + b$" before getting to the final value. So the partial derivative of this node and two input nodes are $\partial c/\partial a = 1$ and $\partial c/\partial b = 1$. Other partial derivatives are labelled in Fig. 4.3.

For the forward-mode differentiation, the calculation flow is from bottom to top. We sum all the path between one input and the final output. For parameter a, we will have

$$\frac{de}{da} = \frac{\partial e}{\partial c} \cdot \frac{\partial c}{\partial a}. \tag{4.14}$$

If we want to get the derivative of de/db, we have to run the simulation once again to obtain the following

$$\frac{de}{db} = \frac{\partial e}{\partial c} \cdot \frac{\partial c}{\partial b} + \frac{\partial e}{\partial d} \cdot \frac{\partial d}{\partial b}. \tag{4.15}$$

. The total simulation time is proportional to the input amount of parameters.

For the backward-mode differentiation, we start from the field value and sum up all the path down to the input parameters. In only one computation, we can get all the derivatives in Eq. 4.14 and Eq. 4.15.

The example I used here only has two parameters, so the time effect is not huge. If the amount of parameters is increasingly large, reverse-mode differentiation will play an important role in saving computation time.

To ease the understanding, let's assume the loss function in Ch. 4.2.1 is uncorrelated with input parameters $\vec{p}$. The total derivative in Eq. 4.9 could be written as

$$\frac{dg}{d\vec{p}} = -\frac{\partial g}{\partial \vec{x}} \cdot (\frac{\partial f}{\partial \vec{x}})^{-1} \cdot \frac{\partial f}{\partial \vec{p}} = -\frac{\partial g}{\partial \vec{x}} \cdot \frac{\partial \vec{x}}{\partial f} \cdot \frac{\partial f}{\partial \vec{p}} = -\lambda^{\intercal} \cdot \frac{\partial f}{\partial \vec{p}}. \tag{4.16}$$

The above Eq. 4.16 is exactly the same as chain rule used in backpropagation in machine learning model training. Mathematically, adjoint method and backpropagation is equivalent. The adjoint equation as in Eq. 4.10 is just the numerical implementation to calculate the derivative by using chain rule.

4.2.3 Inverse design for 3D photonic device

This section presents the inverse design principle for 3D photonic device. There are many literature focusing on adjoint optimization [123] and shape calculus [144]. All of them try to find the sensitivity of merit function with respect to each voxel in the design region within only two simulations. Then the pixel that induces the largest derivative is selected and added to the structure. Next we run a new round of simulations to obtain the new derivatives. The iteration loops until we achieve an optimal design. We will see that the shape calculus [144] used to calculate the gradient is actually the same as adjoint method in Ch. 4.2.2.

Mathematically, the inverse design problem is actually a partial differential equation (PDE) problem, which is

$$\begin{aligned} \max_{\overleftrightarrow{\epsilon},\overleftrightarrow{\mu}} \quad & F(\mathbf{E},\mathbf{H},\omega) \\ \textit{subject to} \quad & \nabla \cdot \overleftrightarrow{\epsilon}\mathbf{E} = \rho \\ & \nabla \cdot \overleftrightarrow{\mu}\mathbf{H} = 0 \\ & \nabla \times \mathbf{E} = -j\omega\overleftrightarrow{\mu}\mathbf{H} \\ & \nabla \times \mathbf{H} = \mathbf{J} + j\omega\overleftrightarrow{\epsilon}\mathbf{E}, \end{aligned} \tag{4.17}$$

where the tensor permittivity $\overleftrightarrow{\epsilon}$ and tensor permeability $\overleftrightarrow{\mu}$ represents material distribution. When solving the PDE problem in Eq. 4.17, we try to find the optimal structure (material distribution) to maximize the predefined merit function.

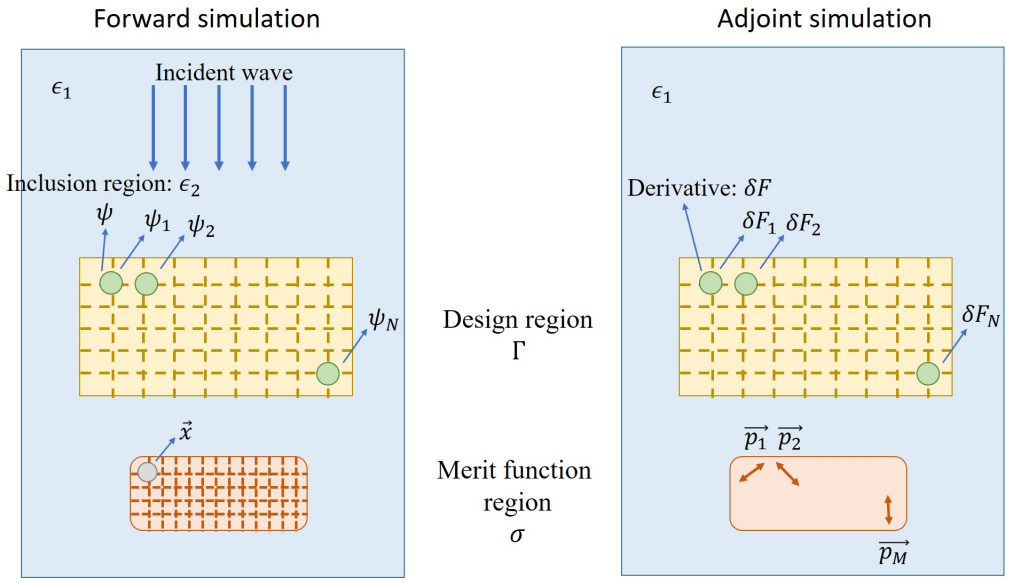

Figure 4.4: Demonstration of calculating the derivative of merit function with respect to each voxel within only two simulations. The background permittivity is ϵ_1. The yellow box represents the design region Γ, while the red box denotes the merit function region σ. In the forward simulation, an incident wave is impinged onto the structure. And we simulate the wave propagation through Maxwell equation to get merit function. Then in the adjoint simulation, each voxel in merit function region is replaced by an adjoint electric and magnetic dipole. We run the adjoint simulation, combined with the electric field data from the forward simulation, and we can get the derivative at each voxel. The voxel inclusion that induces the maximum gradient is selected and added to the design region. Then iterates the next loop to search for the optimal inclusion. In the end, an optimal structure is achieved.

Let's assume the merit function is defined as

$$F = \int_\sigma f(\mathbf{E}(\vec{x}), \mathbf{H}(\vec{x}))d^3\vec{x}, \tag{4.18}$$

where F and f should be a scalar to ease the calculation. And σ denotes the region where field value of all voxels are taken as part of the merit function.

The electric field $\mathbf{E}$ and magnetic field $\mathbf{H}$ are usually a complex number. The derivatives with respect to the real part and imaginary part are independent of each other. To ease the mathematical deduction, we will use the conjugate to distinguish the two different derivatives. Therefore, we have four variables $\mathbf{E}$, $\overline{\mathbf{E}}$, $\mathbf{H}$, $\overline{\mathbf{H}}$, where the top bar denotes the conjugate value.

If we add a small inclusion in the design region Γ, the merit function will have a deviation of

$$\delta F = \int_\sigma \Big(\frac{\partial f}{\partial \mathbf{E}}(\vec{x}) \cdot \delta\mathbf{E}(\vec{x}) + \frac{\partial f}{\partial \overline{\mathbf{E}}}(\vec{x}) \cdot \delta\overline{\mathbf{E}}(\vec{x}) + \frac{\partial f}{\partial \mathbf{H}}(\vec{x}) \cdot \delta\mathbf{H}(\vec{x}) + \frac{\partial f}{\partial \overline{\mathbf{H}}}(\vec{x}) \cdot \delta\overline{\mathbf{H}}(\vec{x}) \Big) d^3\vec{x}. \tag{4.19}$$

Since $\partial f/\partial \mathbf{E}$ is the conjugate of $\partial f/\partial \overline{\mathbf{E}}$ (same as $\partial f/\partial \mathbf{H}$ and $\partial f/\partial \overline{\mathbf{H}}$), we could simplify Eq. 4.19 as

$$\delta F = 2Re\Big[\int_\sigma d^3\vec{x}\big(\frac{\partial f}{\partial \mathbf{E}}(\vec{x})\cdot\delta\mathbf{E}(\vec{x}) + \frac{\partial f}{\partial \mathbf{H}}(\vec{x})\cdot\delta\mathbf{H}(\vec{x})\big)\Big], \tag{4.20}$$

where Re means the real part.

To start with, we run the 1^{st} simulation, which is the forward simulation in Fig. 4.4. We can get the electric field information, $\mathbf{E}^{old}(\vec{x})$, across the whole space.

In the design region Γ, if we add a small inclusion ψ at $\vec{x'}$, (which should be considered as a 3D sphere,) it will be polarized by the original electric field $\mathbf{E}^{old}(\vec{x})$. Then the polarization will affect the electric field, and vice versa. In the ending balanced state, the polarization should be

$$P^{ind}(\vec{x'}) = \overline{\overline{\alpha}}\mathbf{E}^{old}(\vec{x}), \tag{4.21}$$

where according to Claussiu-Mossotti relationship[2],

$$\overline{\overline{\alpha}} = 3\epsilon_0\big(\frac{\epsilon_2-\epsilon_1}{2\epsilon_2+\epsilon_1}\big), \tag{4.22}$$

where ϵ_2 denotes the permittivity of the inclusion material.

Next,we have the Green function to calculate the electric field from a dipole as

$$G^{EP}(\vec{x},\vec{x'}), \tag{4.23}$$

where $\vec{x'}$ is the position of the electric dipole, and $\vec{x}$ is the position of the electric field we want to calculate. For the magnetic field calculation, we have it written as follows

$$G^{HP}(\vec{x},\vec{x'}), \tag{4.24}$$

where $\vec{x}$ is the position of the magnetic field.

The induced electric dipole will cause a deviation in electric field as

$$\delta\mathbf{E}(\vec{x}) = \int_\psi G^{EP}(\vec{x},\vec{x'})P^{ind}(\vec{x'})d^3\vec{x'}, \tag{4.25}$$

and in magnetic field as

$$\delta\mathbf{H}(\vec{x}) = \int_\psi G^{HP}(\vec{x},\vec{x'})P^{ind}(\vec{x'})d^3\vec{x'}. \tag{4.26}$$

Therefore, we can rewrite the variation of merit function in Eq. 4.20 as

$$\delta F = 2Re\Big[\int_\sigma d^3\vec{x}\big(\frac{\partial f}{\partial \mathbf{E}}(\vec{x})\cdot G^{EP}(\vec{x},\vec{x'})P^{ind}(\vec{x'}) + \frac{\partial f}{\partial \mathbf{H}}(\vec{x})\cdot G^{HP}(\vec{x},\vec{x'})P^{ind}(\vec{x'})\big)\Big]. \tag{4.27}$$

Due to the symmetry [144], the Green function could be stated as

$$G^{EP}_{ij}(\vec{x},\vec{x'}) = G^{EP}_{ji}(\vec{x'},\vec{x}) \tag{4.28}$$

$$G_{ij}^{HP}(\vec{x},\vec{x'}) = -G_{ji}^{HP}(\vec{x'},\vec{x}), \tag{4.29}$$

where Einstein notation is used.

Then, the variation of merit function could be simplified as

$$\delta F = 2Re\Big[\int_{\psi} P^{ind}(\vec{x'})d^3\vec{x'}\int_{\sigma} d^3\vec{x}\Big(G_{ji}^{EP}(\vec{x'},\vec{x})\frac{\partial f}{\partial \mathbf{E}_i}(\vec{x}) - G_{ji}^{HP}(\vec{x'},\vec{x})\frac{\partial f}{\partial \mathbf{H}_i}(\vec{x})\Big)\Big]. \tag{4.30}$$

We can assign the adjoint field a value of

$$\mathbf{E}^A(\vec{x'}) = \int_{\sigma} d^3\vec{x}\Big(G_{ji}^{EP}(\vec{x'},\vec{x})\frac{\partial f}{\partial \mathbf{E}_i}(\vec{x}) - G_{ji}^{HP}(\vec{x'},\vec{x})\frac{\partial f}{\partial \mathbf{H}_i}(\vec{x})\Big), \tag{4.31}$$

which can be considered as a radiating result from an electric dipole with an amplitude of $\partial f/\partial \mathbf{E}_i(\vec{x})$ and a magnetic dipole with an amplitude of $-1/\mu_0 \cdot \partial f/\partial \mathbf{H}_i(\vec{x})$.

The variation in merit function in Eq. 4.30 will become

$$\delta F = 2Re\int_{\psi} P^{ind}(\vec{x'})\mathbf{E}^A(\vec{x'})d^3\vec{x'}. \tag{4.32}$$

When the added inclusion is small enough, we can consider the adjoint field is constant across the inclusion. The variation in Eq. 4.32 can be stated as

$$\delta F = 2V \cdot Re\Big(P^{ind}(\vec{x'})\mathbf{E}^A(\vec{x'})\Big), \tag{4.33}$$

where V stands for the volume of the sphere inclusion.

If we run 2^{nd} adjoint simulation in Fig. 4.4 with electric dipoles and magnetic dipoles referenced from Eq. 4.31, we can get the adjoint field $\mathbf{E}^A(\vec{x'})$ across the space, thus obtaining the variation in Eq. 4.33.

In sum, only two simulations are needed to obtain the variation of the merit function. In the 2^{nd} adjoint simulation, the electric dipoles and magnetic dipoles are calculated from the 1^{st} forward simulation. Combing the electric field of the two simulations, we can calculate how much each inclusion in design region can affect the merit function. When we have the performance map across the design region, we can select the inclusion to maximize our merit function. The optimization process loops until it reaches the stopping point.

4.3 Theory for optical coherence tomography (OCT)

Optical coherence tomography (OCT) has been widely applied in deep tissue imaging [145], disease diagnosis [146], etc., ever since its invention in the 1990s [147]. To date, it is continually being enhanced in many aspects, including resolution [148], imaging depth [149][150], and imaging speed [129].

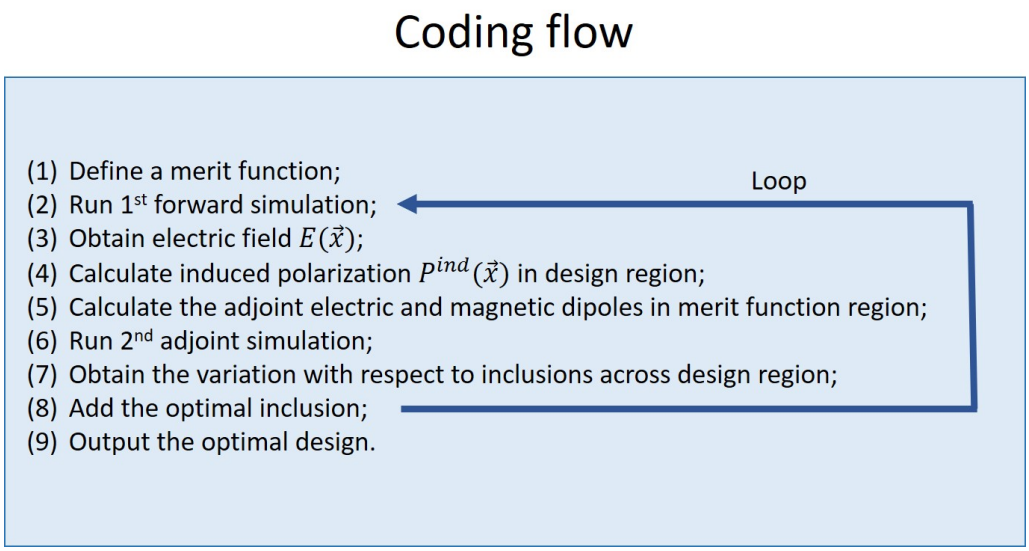

Figure 4.5: The coding flow of the adjoint optimization. Only two simulations are needed to obtain the variation of merit function. Then gradient descent algorithm is applied to find the optimal design.

The theory of OCT is comprehensive in literature [151]. Here I will briefly present the process to obtain an image from an OCT modality. The basic schematic setup of OCT is demonstrated in Fig. 4.6.

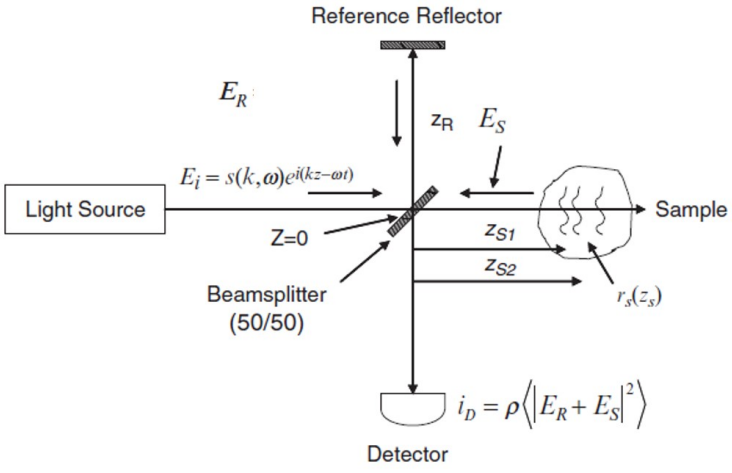

Figure 4.6: The schematic demonstration of an OCT Michelson interferometer [151].

From the light source, we firstly have a low coherent incident beam

$$\mathbf{E}_i = S(k,\nu)e^{i(kz-\omega t)}, \tag{4.34}$$

where $S(k,\nu)$ is the electric field amplitude as a function of the wavenumber $k = 2\pi/\lambda$

and angular frequency $\omega = 2\pi\nu$. Due to dispersion, we define the refractive index as $n(\lambda)$. The relationship between the wavenumber and angular frequency can be formulated as $c/n(\lambda) = \lambda\nu$.

When the incident beam reaches the beamsplitter (, where we set the distance as $z = 0$), it will be splitted under a $50:50$ ratio into the respective reference arm and sample arm. For the reference arm, we set the intensity reflectivity as R_R, the electric field reflectivity being $r_R = \sqrt{R_R}$. The electric field travelling back from the reference reflector will be

$$E_R = \frac{E_i}{\sqrt{2}} S(k,\nu) \cdot r_R \cdot e^{i\cdot 2kz_R}, \tag{4.35}$$

where z_R denotes the distance between the beamsplitter and the reference reflector.

For the sample, we assume the tissue has different electric field reflectivity r_S across depth. We can express the tissue as a layered structure with a reflectivity of

$$r_S(Z_S) = \sum_{n=1}^{N} r_{S_n} \delta(Z_S - Z_{S_n}), \tag{4.36}$$

where each reflectivity r_{S_i} corresponds to a pathlength of Z_{S_i}. The imaging of a sample issue is actually a problem of solving the reflectivity distribution r_{S_n}.

Therefore, the electric field at the beamsplitter reflected from the sample will be

$$E_S = \frac{E_i}{\sqrt{2}} \sum_{n=1}^{N} r_{S_n} e^{i\cdot 2kZ_{S_n}}. \tag{4.37}$$

When the reference beam in Eq. 4.35 and the sample beam in Eq. 4.37 pass the beamsplitter and interfere at the photodetector, the photocurrent will be

$$I_D(k,\omega) = \frac{\rho}{2} |E_R + E_S|^2, \tag{4.38}$$

where ρ represents the responsivity of the photodetector, and the division of 2 is caused by a beam splitting ratio of 50:50. Eq. 4.38 can be detailed as

$$I_D(k,\omega) = \frac{\rho}{2} \left| \frac{E_i}{\sqrt{2}} S(k,\nu) \cdot r_R \cdot e^{i\cdot 2kz_R} + \frac{E_i}{\sqrt{2}} \sum_{n=1}^{N} r_{S_n} e^{i\cdot 2kZ_{S_n}} \right|^2, \tag{4.39}$$

which can be reformulated as

$$\begin{aligned} I_D(k) =& \frac{\rho}{4} \left[S(k) \cdot (R_R + R_{S_1} + R_{S_2} + ... + R_{S_N}) \right] \quad "DC - terms" \\ &+ \frac{\rho}{2} \left[S(k) \cdot \sum_{n=1}^{N} \sqrt{R_R R_{S_n}} \cdot cos(2k(Z_R - Z_{S_n})) \right] \quad "cross - correlation" \\ &+ \frac{\rho}{4} \left[S(k) \cdot \sum_{n \neq m=1}^{N} \sqrt{R_{S_n} R_{S_m}} \cdot cos(2k(Z_{S_n} - Z_{S_m})) \right] \quad "auto - correlation". \end{aligned} \tag{4.40}$$

The inverse Fourier transform $\gamma(Z)$ for the electric field amplitude across spectrum $S(k)$ is

$$\gamma(Z) = e^{-z^2\Delta k^2} \overset{F}{\longleftrightarrow} S(k) = \frac{1}{\Delta k\sqrt{\pi}} e^{-[\frac{(k-k_0)}{\Delta k}]^2}. \tag{4.41}$$

If we apply a Fourier transform to the photocurrent in Eq. 4.40, we will get

$$\begin{aligned} i_D(z) =& \frac{\rho}{8}\Big[\cdot(R_R + R_{S_1} + R_{S_2} + ... + R_{S_N})\Big] \quad "DC-terms" \\ &+ \frac{\rho}{4}\Big[\gamma(Z) \otimes \sum_{n=1}^{N} \sqrt{R_R R_{S_n}} \cdot \delta(Z \pm 2(Z_R - Z_{S_n}))\Big] \quad "cross-correlation" \\ &+ \frac{\rho}{8}\Big[\gamma(Z) \otimes \sum_{n\neq m=1}^{N} \sqrt{R_{S_n} R_{S_m}} \cdot \delta(Z \pm 2(Z_{S_n} - Z_{S_m}))\Big] \quad "auto-correlation". \end{aligned} \tag{4.42}$$

Since the delta function has a sifting property, Eq. 4.42 can be simplified as

$$\begin{aligned} i_D(z) =& \frac{\rho}{8}\Big[\gamma(z) \cdot (R_R + R_{S_1} + R_{S_2} + ... + R_{S_N})\Big] \quad "DC-terms" \\ &+ \frac{\rho}{4}\sum_{n=1}^{N} \sqrt{R_R R_{S_n}}\Big[\gamma[2(z_R - z_{S_n})] + \gamma[-2(z_R - z_{S_n})]\Big] \quad "cross-correlation" \\ &+ \frac{\rho}{8}\sum_{n\neq m=1}^{N} \sqrt{R_{S_n} R_{S_m}}\Big[\gamma[2(z_{S_n} - z_{S_m})] + \gamma[-2(z_{S_n} - z_{S_m})]\Big] \quad "auto-correlation". \end{aligned} \tag{4.43}$$

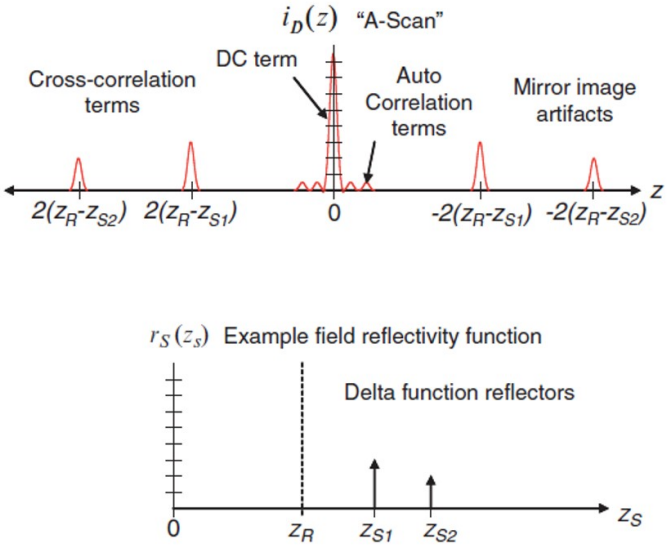

Figure 4.7: Demonstration of the photocurrent in Fourier domain (top). Reconstructed sample reflectivity (bottom)[151].

The "DC-terms" is a constant, which is proportional to the sum of all reflectivities from reference arm and sample arm. The "cross-correlation" term is the desired part

of OCT imaging. We want to solve the square root of sample reflectivities. The "auto-correlation" term is the interference pattern between different sample layers, which will usually cause artifacts in OCT imaging.

The photocurrent in Fourier domain is presented in Fig. 4.7. By selecting the "cross-correlation" term, we can reconstruct the sample reflectivity. Because it only contains the depth information along one axis, it is also called A-scan. When adding a galvo mirror in the OCT setup, we can collect multiple A-scans, which, when combined together, will lead to 2D or 3D tissue imaging results.

4.4 Attenuation-compensation algorithm in OCT imaging

The OCT relies on the reflection of light from structures in different layers to interferometrically reconstruct the volumetric image of the sample. However, light returned from multiple layers suffers from imbalanced attenuation owing to the optical path difference and the inhomogeneous tissue absorption. We report an optimization algorithm to improve signal strength at deep tissue for swept-source (SS)-OCT imaging. This algorithm utilizes the attenuation coefficient of consecutive layers within the sample and combines them to compensate for the signal strength loss from deep tissue. We stacked 170-μm thick cover slides as a standard sample for benchmark testing. The optimized OCT image provides a 30% increase in signal strength in the deep structure compared with the conventional configuration. We applied the method for pearl inspection, whose layered structure demonstrates a great application for our optimized OCT imaging. In contrast to X-ray micro-CT scan and scanning electron microscope (SEM) imaging modalities, the optimized OCT imaging provides great potential for pearl quality inspection. The proposed improvement algorithm for SS-OCT could also be applied to diverse biomedical imaging scenarios, including label-free tissue imaging.

4.4.1 Background

OCT serves as an noninvasive [152] and label-free [153] imaging technique. However, it suffers from image distortion and signal imbalance [154][155], due to, for example, the multiple light scattering from the inhomogeneity of the sample [156]. Usually, such distortion is introduced by the refractive error [157], optical scattering [158], and depth axial attenuation [159]. For refractive error, ray-tracing tools [160] are well-developed and could achieve an accuracy of less than 5×10^{-3} reconstruction error [161]. Meanwhile, a depth-dependent function was proposed to extract the optical scattering coefficient [158], which features to give the heterodyne signal current concerning the depth [162]. In particular, the depth axial attenuation, among all the distortion factors, has attracted substantial attention, since it is frequently encountered in OCT technique and closely associated with refraction and scattering. Based on the single scattering model, an algorithm is proposed to calculate the attenuation coefficient [159]. By a multi-angle

scan, the image distortion from axial attenuation could also be eliminated [163]. As photon propagates randomly in a homogenous medium, a Monto Carlo model can retrieve the signal attenuation and provide calibrated data for compensation of the distortion in OCT images [164]. The attenuation coefficients from existing OCT images could aid to model the axial attenuation to discriminate different tissues [165].

However, most optimization schemes focus on general sample detection. As the performance of one scheme may vary on different sample categories, it is more practical to give a specific solution of a certain sample category of interest. Some works have been demonstrated along this direction with the phantom samples ranging from micro-vasculature [166], slow-rate strains [167], to vascular quantification [168], etc. When it comes to pearls, which is of great artistic value and in need of non-destructive inspection, OCT-based noninvasive detection schemes were also explored, including nacre thickness measurement and nuclei identification [169], defect analysis [170], and treatment investigation [171]. Nevertheless, these schemes adopt the general OCT technique. They could perform better if the OCT is specifically optimized for layered samples like pearls [172].

Here in our work, we report an optimization algorithm called layer-based compensation algorithm (LaB-CA), integrated to the SS-OCT modality, to compensate for the imbalanced attenuation of signal strength from depth. Firstly, a phantom sample, consisting of a stack of cover slides, is investigated to imitate pearls' layered structure. Then, the optimized imaging results of pearls are demonstrated in comparison with X-ray micro-CT images and scanning electron microscope (SEM) images. Our LaB-CA algorithm readily compensates for the layer-dependent attenuation. The results show that LaB-CA is a promising solution for layer-structured samples and specifically, a potential aid for quality inspection and type differentiation in pearls jewelry industry.

4.4.2 Theory and experimental setup

The working principle of our LaB-CA optimized OCT can be simply explained in Fig. 4.8(a), in which the scattering center is supposed to be in each of the multiple layers. As the beam penetrates inside the multiple layers, the light intensity experiences variable attenuation according to the number of layers and the density of the scattering center.

Here, we assume the electric field reflection coefficient of each layer is r_{S_i}, consequent power reflection being $r_{S_i}^2$, where i denotes the order of layers within the sample. In Fig. 4.8(b), if the intensity of the incident beam is unity, it will be reduced to $\prod_{i=1}^{n-1}(1-r_{S_i}^2)$ when it reaches n^{th} layer after attenuation by previous $(n-1)$ layers.

Once reflected, the beam goes through the $(n-1)$ layers again and the intensity is reduced by a factor of $\prod_{i=1}^{n-1}(1-r_{S_i}^2)r_{S_n}^2$ at the detector. The detected intensity for the n^{th} layer is

$$R_{S_n}^2 = \prod_{i=1}^{n-1}(1-r_{S_i}^2)^2 r_{S_n}^2. \tag{4.44}$$

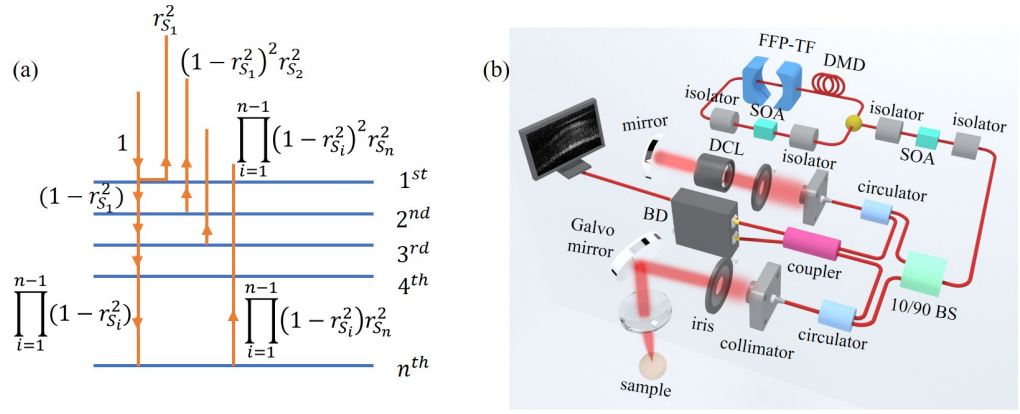

Figure 4.8: (a) Theoretical model: measured reflected signal intensity at different layers in terms of their reflection coefficients. We assume the incident signal intensity is unity. (b) Experimental SS-OCT setup. FFP-TF: fiber Fabry-Perot tunable filter; DMD: dispersion-managed delay; SOA: semiconductor optical amplifier; DCL: dispersion compensation lens; PD: photodetector.

We assume the sample is quasi-uniform within the volume, and conduct the iteration to retrieve the electric field reflection coefficient,

$$r_{S_n} = \sqrt{\frac{R_{S_n}^2}{\prod_{i=1}^{n-1}(1-r_{S_i}^2)^2}}. \tag{4.45}$$

The calibrated reflection coefficient will be applied in subsequent steps to compensate for the imbalanced signal strength. For example, for the next layer, namely, $(n+1)^{th}$ layer, the calculated electric field reflection coefficient would be

$$r_{S_{n+1}} = \sqrt{\frac{R_{S_{n+1}^2}}{\prod_{i=1}^{n}(1-r_{S_i}^2)^2}}, \tag{4.46}$$

with the r_{S_n} being the updated electric field reflection coefficient from Eq. 4.45. In this way, the signal intensity at different detected depth could be balanced.

Fig. 4.8(b) shows our experimental SS-OCT setup based on Michelson-interferometer. The theory of SS-OCT was well discussed previously [151], which utilizes reflected signals from different depths within the sample to reconstruct the image.

We built a Fourier-domain mode-locking (FDML) laser system [173][174]. The fiber Fabry-Perot tunable filter (FFP-TF) (Wolftek, Inc.), powered at 24V, was able to generate an optical spectrum of 80 nm bandwidth centered at 1,500 nm. The semiconductor optical amplifier (SOA) (Inphenix, Inc.) inside and outside the cavity serves as a gain medium and power booster, respectively. The direct output average power of the laser was 3 mW and was split by a 90/10 coupler. The 10% power was used in the reference arm, while the pulses with the remaining 90% power were collimated into the sample

arm. In the reference arm, the beam passes through dispersion compensator (LSM03DC, Thorlabs, Inc.) and retro-reflected by a flat mirror. In the sample arm, the beam was steered by a pair of galvo mirrors (GSV102, Thorlabs, Inc.) and focused by a scanning lens (LSM03, Thorlabs, Inc.) onto the sample. Owing to the axial resolving ability by the optical spectrum, the volumetric imaging only requires the scanning of the beam in two dimensions. The reflected beam from both the sample and the reference arms were collected by the same collimators and split off the main fiber system by the fiber coupler. The interferometric signal was digitalized by the photodetector (PD) for off-line image processing.

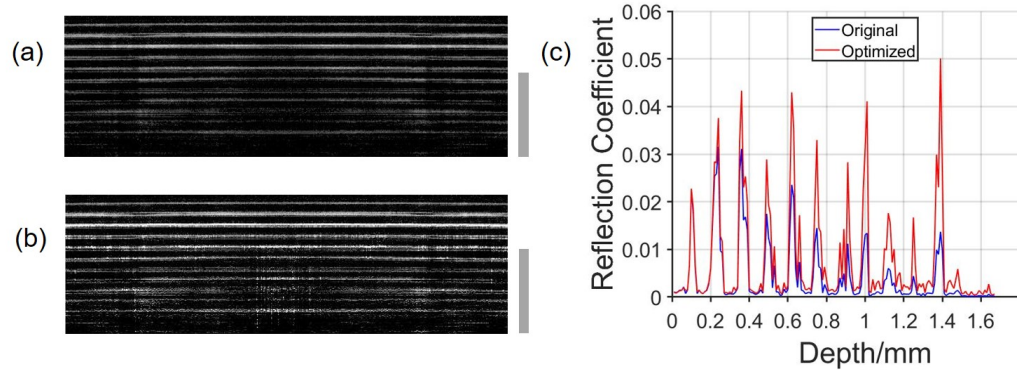

Figure 4.9: (a) Original scanning result of standard cover slides with a thickness of 170 µm. (b) Optimized results after our LaB-CA method. (c) Normalized signal intensity comparison extracted from the imaging results. Scale bar: 1 mm.

To test our algorithm, we run a benchmark experiment to collect the SS-OCT signal retroreflected from the standard sample. Here we prepared a stack of 10 cover slides with a thickness of 170 μm.The original SS-OCT image [Fig. 4.9(a)] suggests a significant intensity decrease concerning the depth. In the presence of our LaB-CA optimization, the SS-OCT image becomes more uniform across the whole imaging depth [Fig. 4.9(b)]. From 3^{rd} layer and onwards (from top to bottom) in Fig. 4.9(b), the signal strength of each layer has increased by more than 30%. Cover slides, composed of glass, have a reflection coefficient of 0.04. To emphasize, no other noise reduction procedure has been applied in our algorithm to avoid any other improvement of the image. Fig. 4.9(c) demonstrates the linear intensity profile along the axial direction before (blue) and after (red) optimization. The intensity becomes more uniform after optimization. Most spikes are close to the reflection coefficient at 0.04, despite the spikes located at about 1.1 mm and 1.3 mm due to their originally weak signal. Since the LaB-CA is to compensate for the weaker signal in deeper tissue rather than fill in the undetected depth beyond, technically the detection depth remains unchanged. Furthermore, other samples, including cellulose (paper) and biological fiber (pork) are investigated to further validate LaB-CA optimization as elaborated in Ch. 4.4.4.

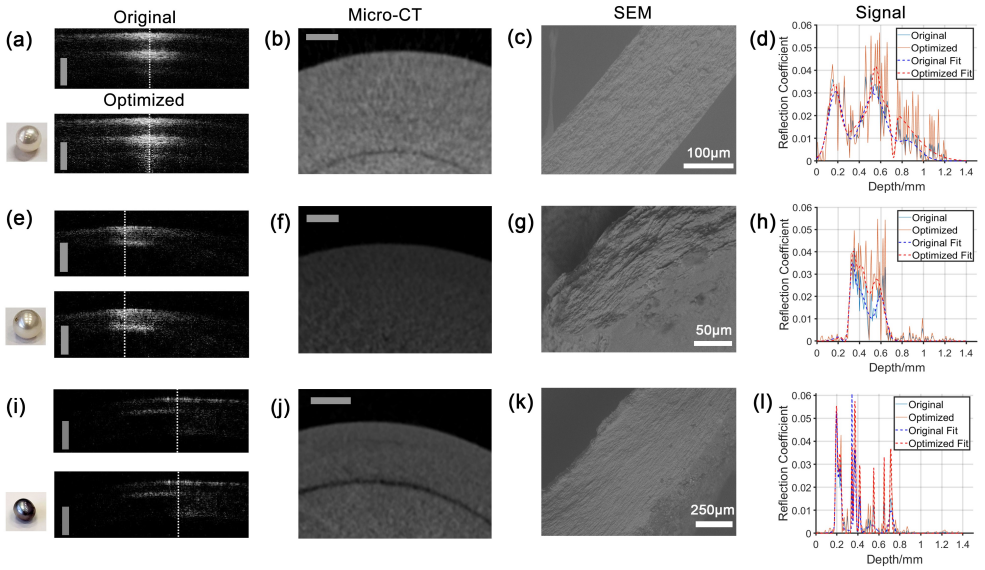

Figure 4.10: The original SS-OCT, LaB-CA optimized OCT, X-ray micro CT, SEM images and signal extraction of different pearls: (a-d) freshwater pearl, (e-h) Akoya pearl (seawater pearl), (i-l) black pearl (seawater pearl). (a, e, i) The OCT images without (top) and with (bottom) optimization. The bottom left corner shows the respective real image of the pearl sample. (b, f, j) The micro-CT images show the morphology of each pearl close to the surface. (c, g, k) SEM images demonstrate the detailed layered structure close to the surface. (d, h, l) The linear OCT signal profile (including original and optimized signal, and Gaussian fit) for each pearl. Scale bar in the first and second column denotes 500 μm. White scale bar on SEM images represents (c) 100 μm, (g) 50 μm, and (k) 250 μm.

4.4.3 Results and discussion

Here we applied the LaB-CA for noninvasive inspection of the pearls shown in Fig. 4.10. The LaB-CA optimized OCT does have an improvement over SS-OCT, which is cross-checked and validated by X-ray micro-CT and SEM images. The pearls we have investigated in Fig. 4.10 are (a-d) freshwater pearl, (e-h) Akoya pearl (seawater pearl), (i-l) black pearl (seawater pearl). For each pearl, we obtained its SS-OCT, LaB-CA optimized OCT, micro-CT (computed tomography), SEM, and signal extraction data, which were listed in the same row as shown in Fig. 4.10. The micro-CT machine (SKYSCAN 1172 X-ray Microtomograph, Bruker Inc.) is set at 10 μm scanning resolution. The SEM's type is Hitachi S-3400N variable pressure SEM.

For the freshwater pearl in Fig. 4.10(a-d), the real image of the sample is shown on the bottom left corner of Fig. 4.10(a). The LaB-CA optimized OCT improves over SS-OCT, in which the signal strength at depth from 0.6 mm to 1.1 mm is stronger than that in original SS-OCT, enabling the visualization of another layer structure beyond the lower nacre. The same pearl sample has also been inspected by micro-CT at the cost of a longer time ($\sim$ hours for a resolution of 10 μm). The micro-CT image close to the edge is shown in Fig. 4.10(b). Although the micro-CT was able to inspect the whole structure

of the large pearl sample ($\sim$ 5 mm in diameter), the resolution is insufficient to observe the details. We also applied the SEM to inspect the nacre structure with high image precision. However, the sectional image requires to break the pearls sample into small pieces. The typical thickness of one nacre layer for the freshwater pearl is roughly 200 μm. We plot the linear intensity profiles from the SS-OCT image with (and without) optimization, and a multi-Gaussian fit to the intensity demonstrating that the FWHM (full width at half maximum) of about 0.2 mm (0.25 mm). The thickness evaluated from the original SS-OCT image over-estimates the nacre thickness. The FWHM retrieved from the optimized image corresponds to the thickness of a single nacre layer, and is consistent with the layer thickness evaluated under SEM. In Fig. 4.10(d), the optimized signal has roughly 8 narrow sharp peaks from 0.45 mm to 0.7 mm in the second Gaussian-like peak. It may give more information on the nacre structure about which the original signal can hardly tell.

The consistency of the layer thickness measurement between the optimized OCT and the SEM has been further corroborated with an Akoya pearl grown in seawater. Fig. 4.10(e) shows the OCT images of the Akoya pearl (photo at the bottom left corner shows the real sample). Similarly, the LaB-CA optimized OCT image shows a clearer and more uniform nacre than the original SS-OCT image. The micro-CT image with uniform signal strength is insufficient to identify the minute structure owing to the lack of resolution ($\sim$ 10 μm). The thickness of the nacre layer is evaluated to be about 100 μm under SEM [Fig. 4.10(g)]. We further extracted the linear signal [Fig. 4.10(h)], which suggests distinct layers located at 0.38 mm, 0.48 mm, and 0.6 mm. So the thickness in FWHM matches well with SEM measurement. As for the original signal, the dips at the mentioned places may be too tiny to be identified as a layer boundary. In each layer, there are nearly 3 peaks at the optimized signal, which is quite limited compared to SEM but gives more information than that of the original one.

We take the black pearl as the second seawater sample [Fig. 4.10(i)], and found that the LaB-CA reinforces slightly on subtle structures displayed in the SS-OCT image. Although we can see the growth ring in the micro-CT image, finer details are impossible to identify. The nacre layer structure has to be imaged under the SEM [Fig. 4.10(k)], which suggests a thickness of the black pearl of around 600 μm. In contrast, the effective signal from the optimized OCT image starts from 0.19 mm to about 0.70 mm [Fig. 4.10(l)]. The consistency of the thickness measurement with LaB-CA optimized OCT and the SEM suggests the feasibility of using the OCT to estimate the layer thickness without breaking the pearl sample itself. It also suggests that the nacre of this black pearl did not fall apart into layers when we cut it. From the Gaussian fit line [Fig. 4.10(i)], it implies that the black pearl may have 8 layers. If it comes to the original signal, the third layer peak located at 0.55 mm and 0.65 mm may be neglected and the original signal ranging from 0.45 mm to 0.60 mm can only give a very limited amount of information due to the weak signal.

Our LaB-CA improves the intensity of original SS-OCT by at least 15% at a deep

nacre layer and gives more accurate results on layer thickness and growth rings, which is further validated by micro-CT and SEM. Based on our LaB-CA optimized results of nacre thickness, growth rings, and reflection coefficient, the jewelry industry may have a better judgment over pearls' type.

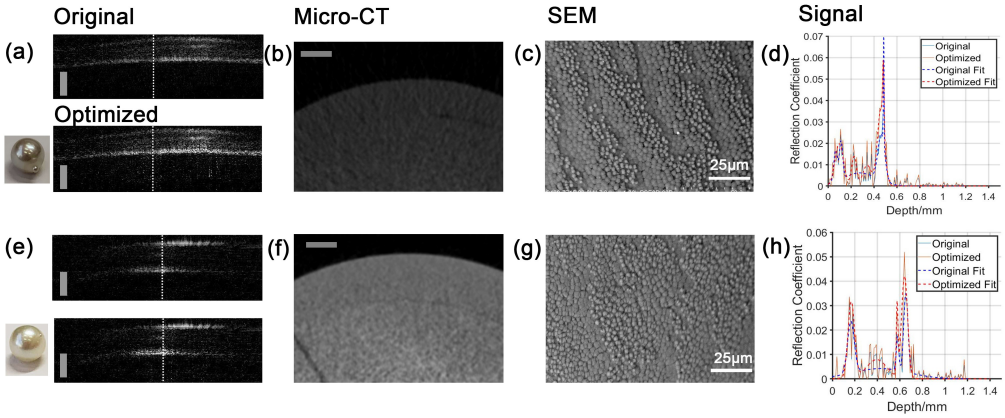

Figure 4.11: Scanning results of two specific pearls, (a-d) bad Akoya pearl, and (e-h) good Akoya pearl. In each row, the data type is the same, viz, SS-OCT data in the first column with the top being original and the bottom being the LaB-CA optimized one, micro-CT scan data in the second column, surface SEM data in the third column, and signal extraction (including signals and their respective Gaussian fit) in the fourth column. The grayscale bar in the first and second column denotes 400 μm, while the white gray bar in (c) and (g) denotes 25 μm.

Secondly, as the quality inspection is of crucial importance in the modern jewelry business, we elect two Akoya pearls (one 'bad' pearl and one 'good' pearl judged by a merchant) for demonstration. Fig. 4.11 shows the imaging results with (a-d) for the bad Akoya pearl and (e-h) for the good Akoya pearl. These two pearls are different from what has been shown in Fig. 4.10.

The micro-CT was able to provide an overall view of the whole nacre, which is consistent with the images shown in Fig. 4.11(a) and 4.11(e), but fails to give more detailed information. For the bad Akoya pearl in Fig. 4.11(d), the signal peaks at 0.22 mm and 0.35 mm of LaB-CA optimized fit is overlooked by the original SS-OCT signal, which is due to the relatively low signal strength in this region. At 0.1-mm depth, the LaB-CA optimized fit has only one peak while the original SS-OCT has two peaks. From SEM images in Fig. 4.10(c) of the surface, the granule of the pearl is sparsely and evenly distributed, which means the layers should have an even distribution along the depth. In that sense, our LaB-CA optimized fit is more reasonable with layer peaks roughly equally spaced along 0 to 0.6 mm than the original SS-OCT.

For the good Akoya pearl in Fig. 4.11(h), the 3 spikes located from 0.3 mm to 0.45 mm have a different impact on the Gaussian fitting. For LaB-CA, a peak is shown, which may suggest a layer, while no peak is plotted on original SS-OCT due to the below-threshold signal strength (0.01 is set as the threshold). Other 3 peaks at 0.15 mm,

0.57 mm, and 0.65 mm, the LaB-CA signal all have at least 20% higher strength than original OCT. From the SEM image in Fig. 4.11(g), the granule distribution is denser, leading to a lower reflection for the layers between upper and lower boundaries. So it can be well explained that the peak at 0.4 mm has an apparent gap compared to the other 3 peaks.

Above all, the LaB-CA optimized OCT may have potential use in quality grading. For the original OCT signal of these two pearls, both show a flat trend from 0.2 mm to 0.4 mm, which is not enough to tell the difference. For LaB-CA optimized OCT, the equally distributed peaks similar to Fig. 4.11(d) may suggest a sparse distribution of the nacre, which is a sign of bad quality. And the 'U'-shape signal, which means the middle signal is much lower than the boundary signals, may suggest the granule is densely distributed and possibly of better quality. Surely, more sample pearls could be selected to set a quality ranking library, to further develop this algorithm.

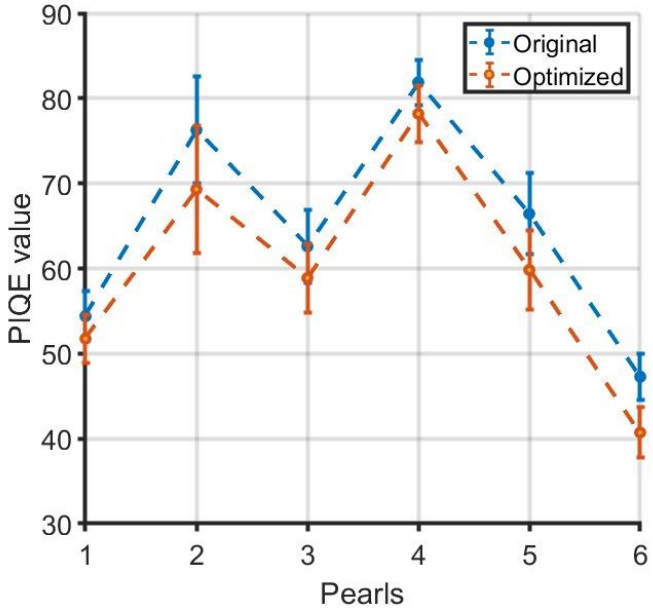

Figure 4.12: PIQE score for the OCT images of the pearl samples, and a lower score represents a higher quality. The blue line denotes original OCT images while the orange line represents the LaB-CA optimized OCT images. In the x-axis, each number stands for different pearls. 1: Freshwater pearl; 2: black pearl; 3-6: different Akoya pearls. (3: the Akoya pearl in Fig. 4.10(e); 4: the bad Akoya pearl in Fig. 4.11(a); 5: the good Akoya pearl in Fig. 4.11(b).

Thirdly, to validate the improvement of our LaB-CA algorithm, we performed the perception-based image quality evaluator (PIQE) to compare the image quality improvement owing to our optimization algorithm. The PIQE score is used to determine image quality without reference [175] by dividing an image into nonoverlapping blocks, extracting the local features to classify whether it is distorted by a preset threshold, and finally providing a PIQE score as a mean of all values from these blocks. In PIQE, a

lower score represents a higher quality. The PIQE score for the pearls of all the sizes we imaged with SS-OCT and processed using the LaB-CA algorithm are shown in Fig. 4.12. In summary, our LaB-CA improves the image quality, or reduces the PIQE score, at about 5 in absolute value, which supports the conclusion that the signal imbalance compensation greatly improves the image quality.

4.4.4 Imaging results for different samples

To validate our algorithm, we apply the LaB-CA optimized OCT to other samples, apart from the pearl samples mainly discussed in the script.

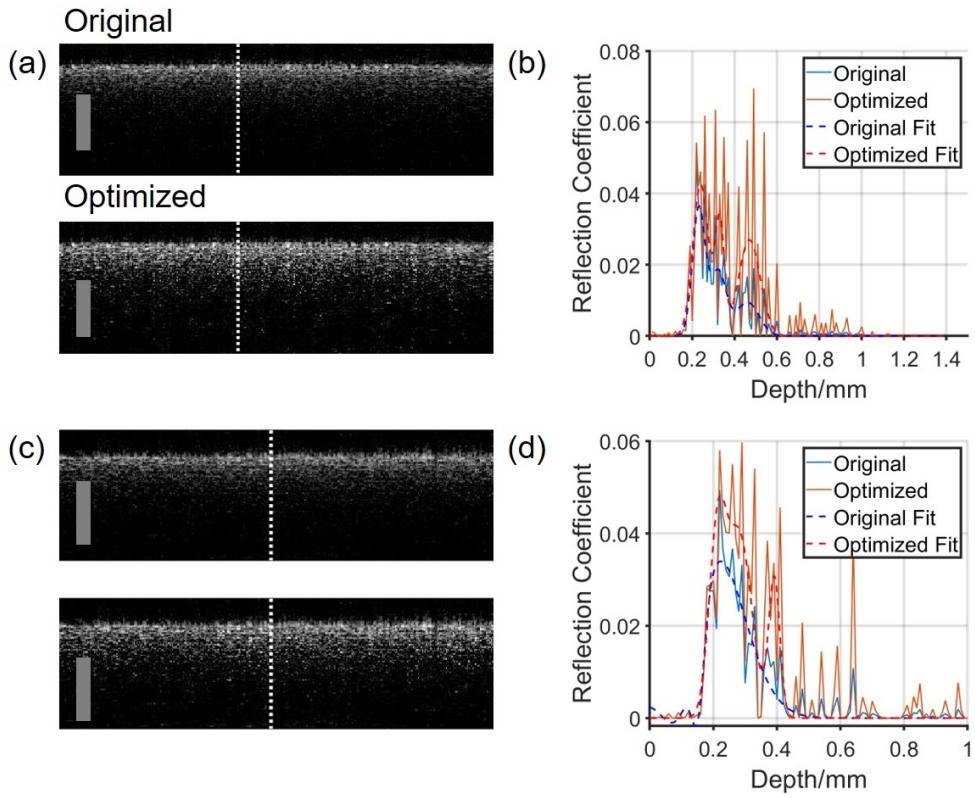

Figure 4.13: OCT scanning results on papers. (a, b) Results for 10 papers as a sample. (c, d) Results for 12 papers as a sample. (a, c) The top image is the original SS-OCT result, and the lower image is our LaB-CA optimized OCT result. (b, d) Extracted signal, along with their Gaussian fit, from the white line in (a) and (c). The grayscale bar is 0.5 mm.

First, cellulose, which is paper in this experiment, is investigated in Fig. 4.13. The thickness of one paper is 0.12 mm, as 70 papers are measured of an 8.50 mm thickness. Fig. 4.13(a) and 4.13(b) present the results of 10 papers, while 4.13(c) and 4.13(d) is about 12 papers. Because of the high absorption of cellulose, the major component of papers, the retrieved signal has decreased after 0.4 mm penetration, which is about 4-paper thickness. For 10-paper results in Fig. 4.13(b), our LaB-CA optimized OCT data's Gaussian fit has 3 peaks, representing that 3 papers are located in the 0.2 mm to 0.5 mm range, while the original SS-OCT data only shows one peak at 0.2 mm, failing to

retrieve the correct information. As the signal's intensity drops dramatically after 0.6 mm, it's acceptable the fourth paper located at 0.6 mm cannot be drawn. For 12-paper results in Fig. 4.13(d), LaB-CA algorithm could identify 2 peaks between 0.2 mm and 0.4 mm, which is correspondent with the paper's thickness measured before. However, the original SS-OCT, still, fails to give useful information and only shows 1 peak at 0.2 mm.

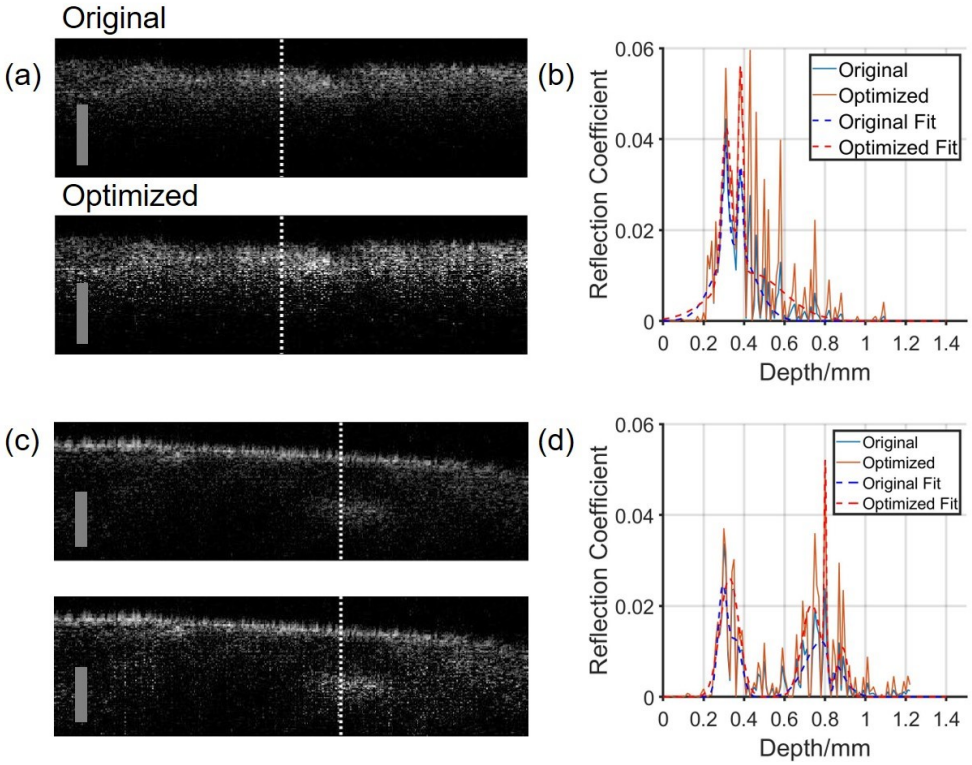

Figure 4.14: OCT scanning results on lean and fatty pork. (a, b) Results for the lean pork. (c, d) Results for the fatty pork. (a, c) The top image is the original SS-OCT result, and the lower image is our LaB-CA optimized OCT result. (b, d) Extracted signal, along with their Gaussian fit, from the white line in (a) and (c). The grayscale bar is 0.5 mm.

Second, biological fiber, which is lean pork in Fig. 4.14(a) and 4.14(b) and fatty pork in Fig. 4.14(c) and 4.14(d) here, is examined under LaB-CA optimization OCT. For the lean pork, the second peak improvement of LaB-CA optimized image at 0.35 mm is quite obvious in Fig. 4.14(b). From 0.4 mm and onwards, the LaB-CA optimized fit may indicate another layer, which is the muscle fiber for this sample, while the original SS-OCT fit just decreases to zero. Besides, for the fatty pork in Fig. 4.14(d), the LaB-CA optimized OCT indicates a layer at 0.72 mm while the original SS-OCT fails. For the peak located at 0.3 mm and 0.9 mm, the optimized fit has a nearly 5% and 21% intensity increase respectively compared to original SS-OCT.

4.4.5 Summary

In conclusion, we reported a LaB-CA algorithm incorporated into SS-OCT to optimize the signal intensity from deeper tissue and displayed its use in pearl imaging to differentiate their types and evaluate their quality. The LaB-CA algorithm considers each layer's scattering and attenuation, then compensates the loss to the deeper layer. It is suitable for those layer-structured samples, e.g., pearls in particular. The benchmark test with standard cover slide stack using LaB-CA suggested an improvement of 30% in signal intensity from the 3rd layer and onwards compared to original SS-OCT. Then we applied it to pearl imaging and demonstrated good correspondence with X-ray micro-CT and SEM image. We showed that our LaB-CA improves the image quality with PIQE score reduced by nearly 5 in absolute value or 8% in ratio.

In contrast to the X-ray phase imaging and neutron imaging [176], the SS-OCT equipped with our LaB-CA algorithm is readily suitable for the non-invasive and label-free inspection of the pearl. LaB-CA optimization has great potential to reveal the relationship between signal trend versus category and quality. In the future, more pearls could be elected to generate different classifications of category and quality [177] for further experiments, to give a general rule of signal trend and pearls' property, e.g., type, origin, and quality. And existing SS-OCT techniques for pearls investigation [169] could benefit from the LaB-CA optimization for more accurate and quantified results.

4.5 Chip-modality: meta-structure for an extended depth of focus

In optical imaging field, extended depth of focus (EDOF) demonstrates great importance in improving imaging depth and resolution [133]. In this regard, Bessel beam and Airy beam has been frequently adopted as a penetration beam due to their diffraction-free features [134]. However, such features come with high-energy side lobes, which will cause background noise and photobleaching [134]. Therefore, the community spends huge efforts in designing different optical devices, e.g. metasurface [135], phase mask [128], to achieve an EDOF with reduced side lobes. However, the forward process, e.g. metasurface geometry sweeping, phase mask designing, may pose high computation burden [134] but lead to a decreased focusing efficiency [135]. Recently, some inverse design methods have been applied in metasurface designing [136]. However, the research on inversely designing a phase mask for an EDOF is very limited.

Here, we present an adjoint optimization method to deliver a phase mask so as to achieve an EDOF. The phase mask could be easily accomplished through a spatial light modulator (SLM) or metasurface. The EDOF will empower optical imaging with a higher resolution and signal-to-noise ratio (SNR) compared to Bessel beam.

Fig. 4.15 presents our schematic setup. The size of the metasurface is 300×300 μm^2. The wavelength is 532 nm. Each pixel has a size of 250×250 nm^2. For a given

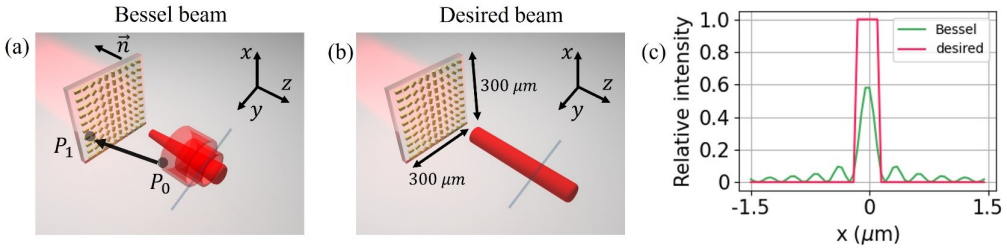

Figure 4.15: Schematic demonstration of the beam profile. (a) Bessel beam intensity distribution. (b) Desired beam intensity distribution without side lobes. (c) Intensity at a cross section.

phase mask $U(\vec{r})$, we can have its far-field radiation $U(P_0)$ through Fresnel-Kirchhoff formula [104]

$$U(P_0) = \frac{1}{j \cdot 2\lambda} \int \frac{exp(jkr_{01})}{r_{01}} \cdot U(\vec{r}) \cdot cos(\vec{n}, \vec{r}_{01}) dS, \tag{4.47}$$

where $\vec{r}_{01}$ points from the field to the phase mask, λ is the wavelength and k is the wavenumber.

It can be written in a nonlinear matrix form

$$g(v, p) = A_{nl}(p)v - b(p) = 0, \tag{4.48}$$

where v is the field value, and p is the tunable phase modulation at each pixel.

Here, we initialize the phase profile $p(x, y)$ for Bessel beam generation. Then, we use the adjoint optimization to update the phase pattern. We firstly propose a desired focus field I_0. The objective function we want to minimize is

$$f(v, p) = \sum_{x,z} abs(I_0 - U), \tag{4.49}$$

where U is the far-field intensity distribution during iteration. For each pixel on the phase mask, we apply a gradient descent updating rule, which is

$$p_{i+1} := p_i - \frac{df}{dp_i}. \tag{4.50}$$

The gradient on each pixel can be obtained through only two adjoint simulation. The adjoint simulation is solving the following equation of

$$(\frac{\partial g}{\partial v})^\intercal \cdot \lambda = -(\frac{\partial f}{\partial v})^\intercal. \tag{4.51}$$

The gradient can be readily calculated from this adjoint simulation as

$$\frac{df}{dp} = \frac{\partial f}{\partial p} + \lambda^\intercal \cdot \frac{\partial g}{\partial p}. \tag{4.52}$$

In the end, we output the phase mask when the objective function is lower than threshold.

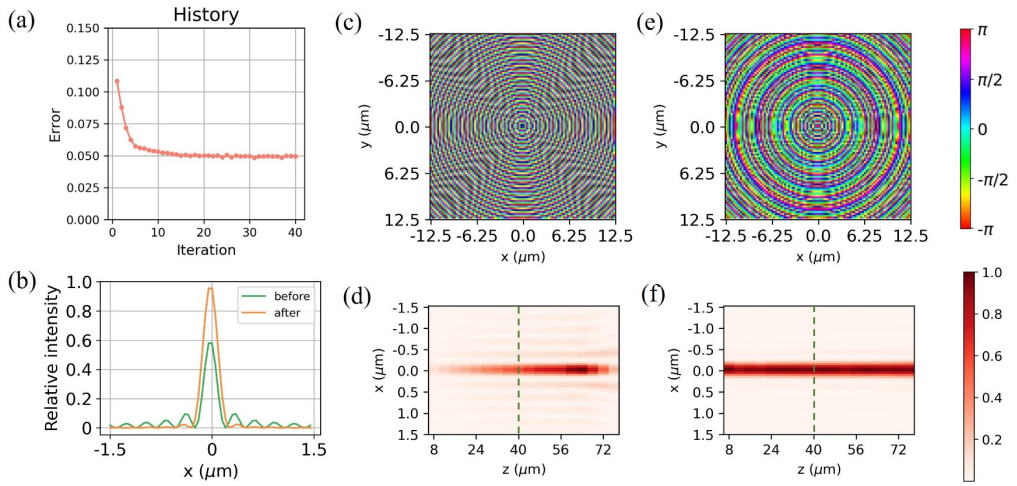

Figure 4.16: (a) Optimization history. (b) Intensity distribution at z-axis distance of 40 µm displayed in (d) and (f) before and after optimization. (c) Initial phase mask. (d) Initial intensity distribution. (e) Optimized phase mask. (f) Optimized intensity distribution.

The optimization result is shown in Fig. 4.16. After optimization, the far-field focus has been greatly improved with a reduced side lobe in Fig. 4.16(e). In Fig. 4.16(a), the error is defined as the deviation from the desired intensity averaged across pixels. During the iteration, the error has been lowered from 10.84% to 4.90%. Once optimized, the phase mask can be uploaded to SLM or fabricated into a metasurface without further calculation. It shows a great potential in an enhanced biomedical tissue imaging.

4.6 Fiber-modality: meta-structure for an extended depth of focus

For chip-modality meta-structure discussed in Ch. 4.5, it will be very easy to integrate a predesigned chip into an existing optical setup so as to achieve an EDOF for an enhanced imaging. However, such chip-based optical setup may still be too bulky for blood vessel or capillary imaging.

To assist vascular disease diagnosis and surgery, the community begins to pay attention to ultrasmall optical imaging solutions [178]. These ultrasmall endoscope systems pave the way in vessel imaging with least destruction.

Here, I present a meta-OCT as in Fig.4.17, whose probe is a single-mode fiber (SMF) meticulously designed to achieve an EDOF. Conventional OCT usually uses a mirror in the sample arm to reflect beam into the sample. The imaging depth is limited by the Rayleigh range of the beam. I propose an inverse design algorithm based on adjoint optimization in Ch. 4.2 to design meta-structure on the SMF. The SMF is cleaved with

a 40° to realize total reflection. If we fabricate the meta-surface on top of the facet, the refected beam will have an EDOF which will increase the imaging depth.

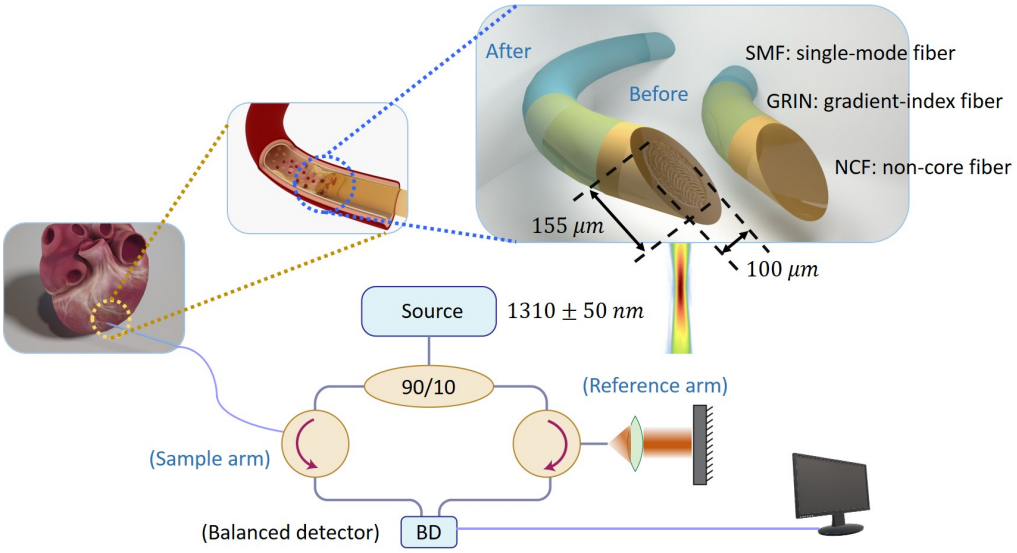

Figure 4.17: Schematic figure for meta-OCT. The OCT setup is well explained in Ch. 4.3. The probe of the sample arm is an SMF with a specially designed metastructure on its facet to achieve an EDOF.

4.6.1 Meta-probe structure

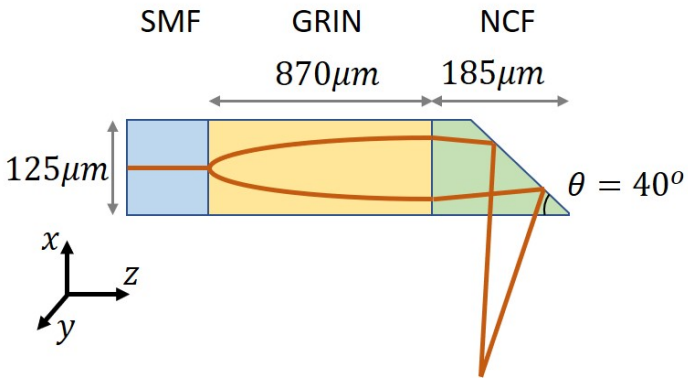

Figure 4.18: Schematic figure of the meta-probe of the sample arm in meta-OCT setup in Fig. 4.17. It consists of SMF, GRIN, and NCF. The beam is directed by SMF, then expanded by GRIN, and in the last reflected by NCF into the tissue sample.

The structure of meta-probe is presented in Fig. 4.18. It consists of SMF, GRIN, and NCF. The beam is directed by SMF, then expanded by GRIN, and in the last reflected by NCF into the tissue sample.

The SMF has a diameter of 125 μm. And it is integrated with a GRIN fiber with a length of 870 μm. And another 185-μm NCF is attached to the end of GRIN. The refractive index profile of GRIN at different radius is measured by a refractive index

profiler (S14, Photon Kinetics, Inc) at 1310 nm and fitted as [179]

$$n(r) = n_g \cdot (1 - \frac{g^2 r^2}{2}), \tag{4.53}$$

where g is 1.8 mm^{-1}, and n_g is 1.4538, and r represents the radius in the unit of micrometer. The GRIN fiber's core has a diameter of 100 μm, while the cladding has an outer diameter of 125 μm. The refractive index for NCF is $n_{NCF} = 1.4470$ at 1310 nm.

The GRIN fiber, in the meta-probe, serves as a beam expander. Here I will present the output beam radius calculation, which also explains the length of 870 μm for GRIN fiber is an optimal choice.

The GRIN fiber can be considered as an inhomogeneous medium. The trajectory of light in such medium is described by Fermat's principle [180]

$$\delta \int_A^B n(\vec{r}) ds = 0, \tag{4.54}$$

where $n(\vec{r})$ represents the refractive index distribution, and ds is the differential length which will be integrated from point A to B.

Using the calculus of variation, Eq. 4.54 can be expressed in a ray equation [180] as

$$\frac{d}{ds}\left(\mathbf{n} \cdot \frac{d\vec{r}}{ds}\right) = \nabla \mathbf{n}, \tag{4.55}$$

where s is the distance along the trajectory, and $\vec{r}$ represents the position of the ray, and $\mathbf{n}$ denotes the refractive index distribution, and gradient ∇ means $(\partial/\partial x, \partial/\partial y, \partial/\partial z)$.

Given the GRIN fiber is cylindrically symmetric, we can simplify the ray equation Eq. 4.55 on the x-z plane, which will lead to

$$\frac{d}{ds}\left(n \cdot \frac{dx}{ds}\right) = \frac{\partial n}{\partial x}, \tag{4.56}$$

where

$$ds = dz\sqrt{1 + \left(\frac{dx}{dz}\right)^2}. \tag{4.57}$$

If the trajectory is a function of $x(z)$, we will, by solving Eq. 4.56, get the following differential equation

$$x''(z) = \frac{1}{n}\left(\frac{\partial n}{\partial x} - \frac{\partial n}{\partial z} \cdot x'(z)\right) \cdot (1 + x'(z)^2). \tag{4.58}$$

Given the refractive index distribution is known from Eq. 4.53, we can rewrite Eq. 4.58 as a 2^{nd} order ordinary differential equation (ODE)

$$x''(z) = -\frac{g^2 x}{1 - \frac{g^2 x^2}{2}} \cdot (1 + x'(z)^2). \tag{4.59}$$

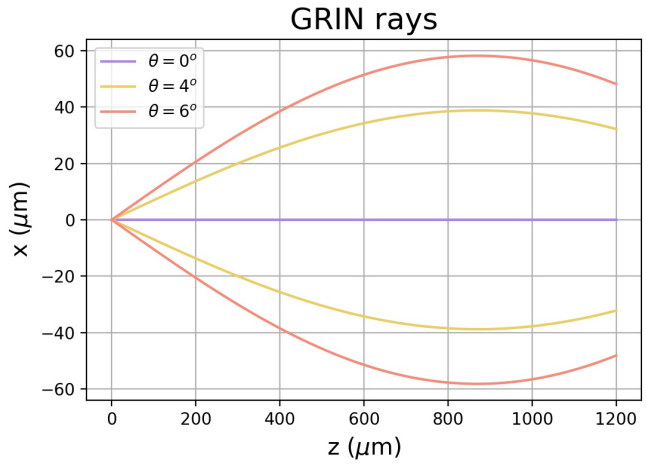

Figure 4.19: Rays trajectory in GRIN fiber in Fig. 4.18. Three different incident angles are plotted, namely, they are 0°, 4°, 6°. The length of GRIN fiber is assumed to be 1200 μm. At 870 μm, the beam output angle is in parallel with the axis.

By solving the ODE in Eq. 4.58, we can get the numerical solution of the ray trajectory $x(z)$, which is demonstrated in Fig. 4.19. The z-axis starts from 0 at the interface between SMF and GRIN fiber. I assume the beam is a point source at the output of SMF. Three incident angles of ray trajectories, which are 0°, 4°, 6°, are plotted as a function of GRIN length. From Fig. 4.19, the output beam ray, at around 870 μm, is parallel to the z-axis.

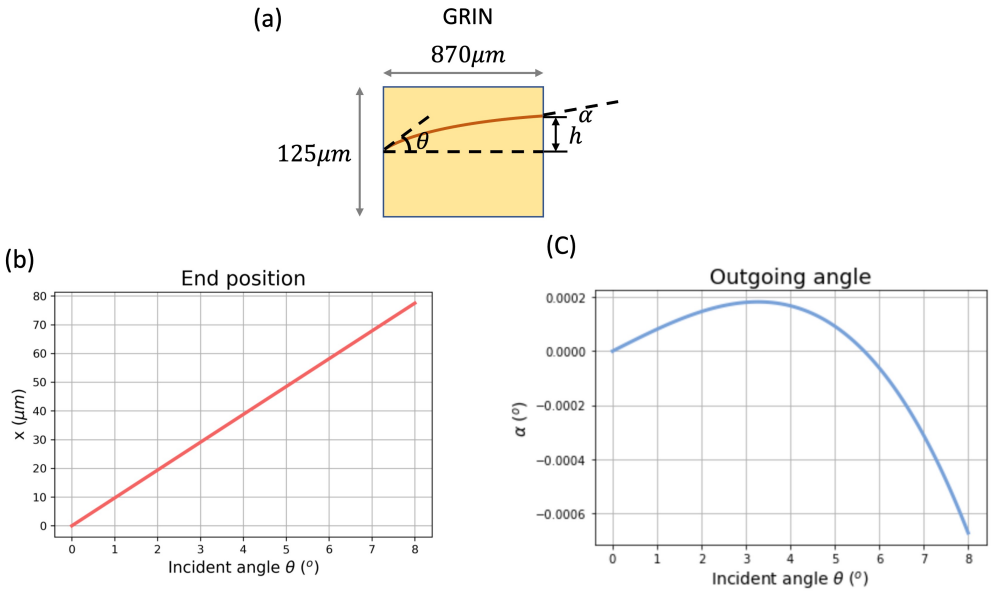

Figure 4.20: The ray trajectory after travelling through the GRIN fiber. (a) Schematic demonstration: for a ray with an incident angle of θ, it will end with an output angle of α at a radius position of h. (b) The end position at the output of GRIN fiber with respect to incident angle. (c) The outgoing angle after the GRIN fiber with respect to the incident angle.

We choose the GRIN fiber length to be 870 μm. Then I plot the ray trajectory output position and output angle with respect to the incident angle in Fig. 4.20. For an incident angle of about 5.2°, the output ray is almost positioned at 50 μm, the margin of GRIN fiber core part. It means for the point source beam, only the cone within 5.2° will be transmitted and expanded toward the end of GRIN fiber. In Fig. 4.20(c), we can conclude for all the rays with incident angle below 5.2°, the output angle is near 0°.

In sum, the GRIN fiber, with a length of 870 μm, functions as a beam expander for 1310-nm source. The expanded beam can be considered as a plane wave.

4.6.2 Geometrical optics solution

This section presents the meta-probe design procedure based on geometrical optics. Firstly we have to calculate the phase map needed at the facet of the meta-probe. Then we will mill structure on the facet to introduce the phase map.

The calculation method is based on Fermat's principle [180]. I define a focal point with a desired working distance as P_3 in Fig. 4.21. The optical path from the starting point, travelling through GRIN and NCF, being reflected by the facet, all the way to the focal point, should be the same across all the rays with different incident angle.

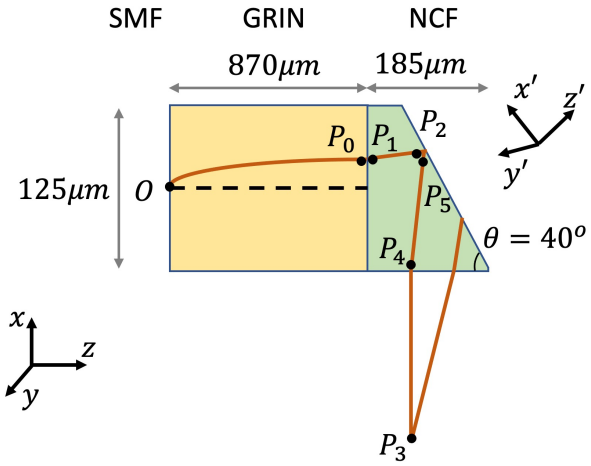

Figure 4.21: Optical path demonstration. O: starting point. P_0, P_1: output point at the end of GRIN. P_2, P_5: the contacting point of the incident ray from P_1 and the end facet. P_4: the contacting point of the reflected beam from P_5 and the periphery of the cylindrical NCF. P_3: desired focal point.

Optical path calculation

The optical path from starting point O to the focal point P_3 can be calculated as

$$L_{OP_3} = L_{OP_2} + L_{P_3P_5}. \tag{4.60}$$

I firstly calculate the optical path from starting point O to the facet P_2 as

$$L_{OP_2} = L_{OP_0} + L_{P_1P_2} = \int_O^{P_0} n(r)ds + n_{NCF} \cdot |P_1P_2|. \tag{4.61}$$

Then I compute the optical path from the focal point P_3 to the facet P_5 as

$$L_{P_3P_5} = n_{air} \cdot |P_3P_4| + n_{NCF} \cdot |P_4P_5|. \tag{4.62}$$

If we want to focus beam into point P_3, all the rays from point O should have the same phase when travelling to point P_3. Therefore, we will have

$$(L_{OP_2} + L_{P_3P_5}) \cdot \frac{2\pi}{\lambda_0} + \phi = L_{shortest} \cdot \frac{2\pi}{\lambda_0}, \tag{4.63}$$

where ϕ is the phase map we want to introduce at the facet, and λ_0 is the working wavelength, and $L_{shortest}$ means the shortest optical path among all the rays.

Therefore, the desired phase mask at the facet should be

$$\phi = L_{shortest} \cdot \frac{2\pi}{\lambda} - (L_{OP_2} + L_{P_3P_5}) \cdot \frac{2\pi}{\lambda_0}. \tag{4.64}$$

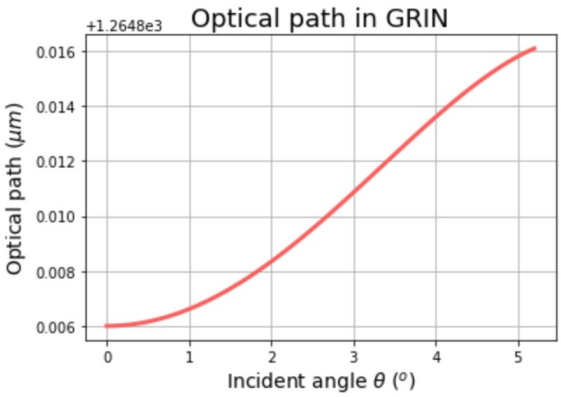

Figure 4.22: The optical path in GRIN fiber with respect to incident angle.

The optical path in GRIN fiber is calculated as

$$L_{OP_0} = \int_O^{P_0} n(r)ds, \tag{4.65}$$

which is plotted in Fig. 4.22.

Then I will calculate the optical path from point P_1 to point P_2 as in Fig. 4.23. We firstly obtain the Cartesian coordinate of P_0 : (x_0, y_0, z_0), which can be expressed in cylindrical form as (r_0, θ_0, z_0). The outgoing angle at P_0 is a function of radius position, which can be mapped as $\alpha_0 = \alpha(r_0)$ referenced from Fig. 4.20. The angle of the beam

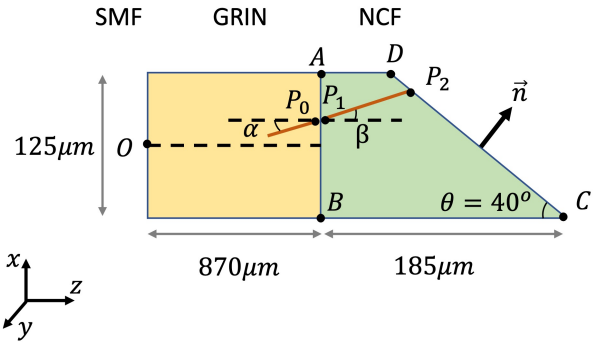

Figure 4.23: The optical path from point P_1 to point P_2. α is the incident angle when the ray reaches the end of GRIN fiber. β is the outgoing angle of the ray in NCF. $\vec{n}$ is the normal vector of the facet CD. (The graphic geometry may not represent the real geometry. Some parts has been enlarged in the figure to increase the clarity. The exact size should refer to the label.)

going out into the NCF can be expressed as

$$\sin\beta_1(r_0) = \frac{n_{GRIN}(r_0)}{n_{NCF}} \cdot \sin\alpha_0(r_0). \tag{4.66}$$

It is easy to write the direction of $\overrightarrow{P_1P_2}$ in Cartesian form

$$\hat{v}_1 = (\sin\beta_1\cos\theta_0, \sin\beta_1\sin\theta_0, \cos\beta_1). \tag{4.67}$$

Let's assume the point P_2 has a coordinate of (x_2, y_2, y_2). We can solve the position by

$$\begin{aligned} \overrightarrow{P_1P_2} &= k_0\hat{v}_1 \\ \overrightarrow{P_2C} \cdot \vec{n} &= 0, \end{aligned} \tag{4.68}$$

where k_0 is a constant.

The optical path of $|P_1P_2|$ can be calculated as

$$L_{P_1P_2} = n_{NCF} \cdot ||P_1P_2||. \tag{4.69}$$

Then, the original point is shifted to the center of $|AB|$ to ease the math. I define a working distance of f, which is the length from focal point P_3 to the center of the facet E. We will have the coordinate of P_3 as $(-f, 0, |BC| - \frac{1}{2}\frac{|AB|}{\tan\theta})$. We assume the ray comes out of point P_3 with a solid angle of (φ, δ), which is defined in the inset of Fig. 4.24. The direction of $\overrightarrow{P_3P_4}$ can be written as

$$\hat{v}_3 = (\cos\varphi, \sin\varphi\cos\delta, \sin\varphi\sin\delta). \tag{4.70}$$

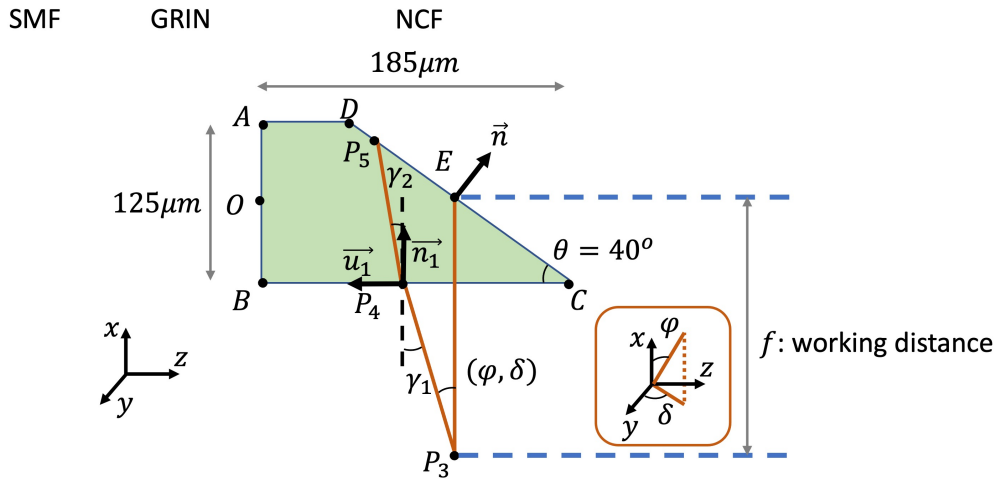

Figure 4.24: Optical path calculation from point P_3 to point P_5. The original point is positioned at the center of AB.

Let's assume the point P_4 has a coordinate of (x_4, y_4, y_4). We can solve the position by

$$\begin{aligned} \overrightarrow{P_3P_4} &= t_0\hat{v}_3 \\ x_4^2 + y_4^2 &= \frac{|AB|^2}{4}, \end{aligned} \tag{4.71}$$

where t_0 is a constant. The optical path of $|P_3P_4|$ can be calculated as

$$L_{P_3P_4} = n_{air} \cdot ||P_3P_4||. \tag{4.72}$$

The direction of $\overrightarrow{P_4P_5}$ can be calculated as

$$\hat{v}_4 = \sin\gamma_2 \cdot \overrightarrow{u_1} + \cos\gamma_2 \cdot \overrightarrow{n_1}, \tag{4.73}$$

where $\overrightarrow{n_1}$ is the normal vector of the cylinder, and $\overrightarrow{u_1}$ is the tangential vector of the cylinder, and γ_2 is derived from Snell law by $\sin\gamma_2 \cdot n_{NCF} = \sin\gamma_1 \cdot n_{air}$. And γ_1 can be calculated by $\cos\gamma_1 = \overrightarrow{n_1} \cdot \hat{v}_3$.

Let's assume the point P_5 has a coordinate of (x_5, y_5, y_5). We can solve the position by

$$\begin{aligned} \overrightarrow{P_4P_5} &= s\hat{v}_4 \\ \overrightarrow{P_5C} \cdot \vec{n} &= 0, \end{aligned} \tag{4.74}$$

where s is a constant. The optical path of $|P_4P_5|$ can be calculated as

$$L_{P_4P_5} = n_{NCF} \cdot ||P_4P_5||. \tag{4.75}$$

In sum, the total optical path from SMF to the focal point is, by combining the results fromEq. 4.65, Eq. 4.69, Eq. 4.72, Eq. 4.75,

$$L_{OP_3} = \int_O^{P_0} n(r)ds + n_{NCF} \cdot ||P_1P_2|| + n_{air} \cdot ||P_3P_4|| + n_{NCF} \cdot ||P_4P_5||. \tag{4.76}$$

Phase mask calculation

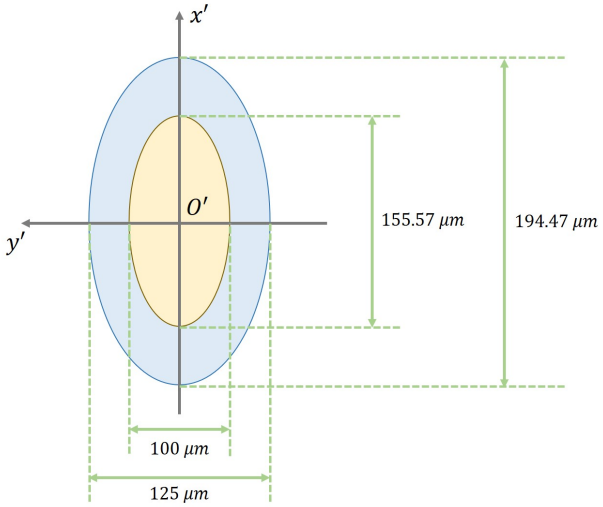

Figure 4.25: Shifted coordinate system to demonstrate the phase profile on the facet. Blue region is the total facet section. Yellow region is the illuminated part from GRIN fiber.

Since the total optical path in the original probe is calculated in Eq. 4.76, we can derive the phase profile we want to introduce on the facet from Eq. 4.64 as

$$\phi = (L_{shortest} - ||L_{OP_3}||) \cdot \frac{2\pi}{\lambda_0}. \tag{4.77}$$

Firstly I shift the coordinate with a $x' - y'$ plane parallel to the facet and a original point O' at the center of facet, which is demonstrated in Fig. 4.21. The new coordinate is also plotted in Fig. 4.25.

We selected two different working distance, namely, they are 1500 μm and 2000 μm. Firstly, we calculate the total optical path for each ray, which is plotted with respect to each ray's position when it reaches the facet in Fig. 4.26(a) for $f = 1500\mu m$ and in Fig. 4.26(c) for $f = 2000\mu m$.

Among all the rays, the shortest optical path is selected as the $L_{shortest}$. The phase mask required on the facet is calculated from Eq. 4.77 for $1500\mu m$ in Fig. 4.26(b) and for $2000\mu m$ in 4.26(d).

The desired phase profile on the facet in Fig. 4.26(b) and Fig. 4.26(d) is continuous, which is impossible to fabricate. So I discretize the phase profile in Fig. 4.27 for Fig.

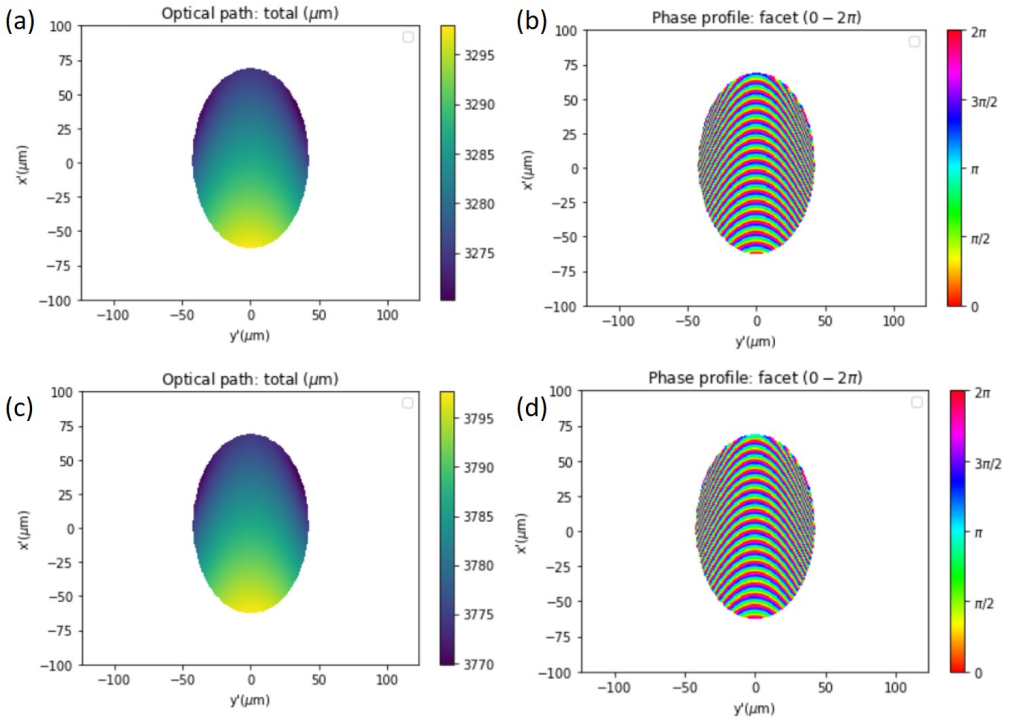

Figure 4.26: The desired phase mask on the facet. When the working distance f is 1500 μm, the total optical path is demonstrated in (a) and the desired phase mask is shown in (b). When the working distance f is 2000 μm, the total optical path is demonstrated in (c) and the desired phase mask is shown in (d). The figures are plotted on a $x' - y'$ plane parallel to the facet as shown in Fig. 4.21. The original point is the geometry center of the elliptical facet. (Here the phase map is calculated based on the GRIN fiber with a length of 100 μm.)

4.26(b) with a working distance of 1500 μm and in Fig. 4.28 for Fig. 4.26(d) with a working distance of 2000 μm.

Structure design

This section presents the methods of introducing a phase mask on the facet, which is shown in Fig. 4.29. On the facet, different depth is fabricated to apply the according phase. This section will provide the relationship between the fabricated depth and the desired phase mask.

In Fig. 4.29(b), no structure is applied. The incident beam will go through the total internal reflection and transmit out of the probe in its side face. The incident angle is $\eta = (90° - \theta) = 50°$. I assume the medium outside the probe is air. The critical angle will be $\theta_C = \arcsin(n_{air}/n_{NCF}) = 43.7°$. Since $50° > \theta_C$, the condition of total internal reflection is satisfied.

In Fig. 4.29(c), a depth is milled onto the facet. It is obvious that beam $\vec{E}_2$ travels longer than beam $\vec{E}_1$, which imposes a phase delay. The optical path of beam $\vec{E}_2$ is

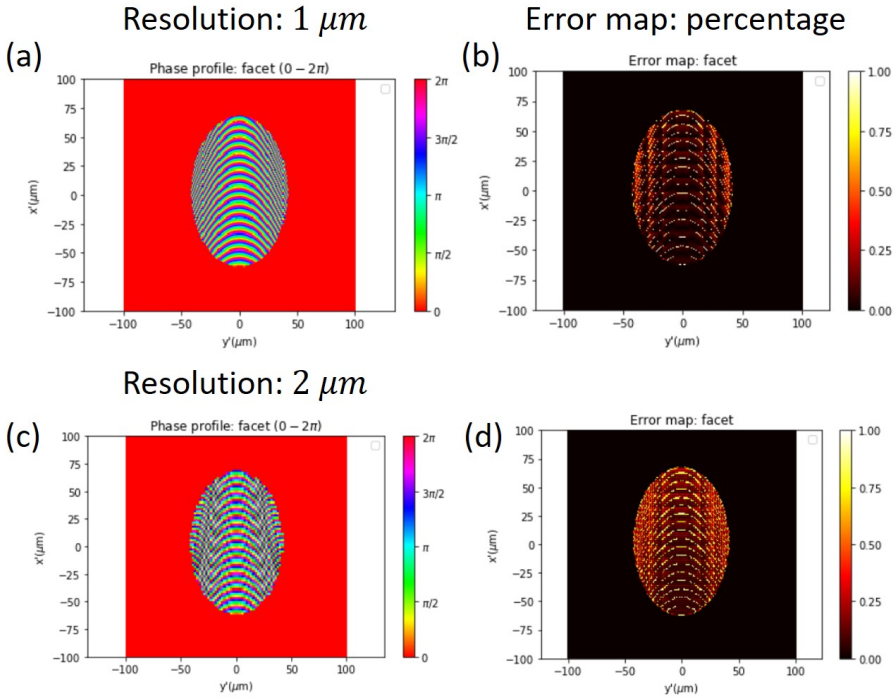

Figure 4.27: Discretized phase profile for the original phase mask in Fig. 4.26(b). (a) The phase profile for 1-μm resolution. I divide the whole phase map into a meshgrid with a lattice size of $1 \times 1 \mu m^2$. For each lattice, I take the averaged phase as the final phase profile. (b) The deviation error of the discretized phase profile in (a) and the original phase profile. (c) The phase profile for 2-μm resolution. (d) The deviation error of the discretized phase profile in (c) and the original phase profile.

longer than beam $\vec{E}_1$ with a value of

$$L_{\Delta\vec{E}_2\vec{E}_1} = \frac{d}{\cos\eta}\Big(1 - \sin\left(2\eta - 90^\circ\right)\Big) \cdot n_{NCF}, \tag{4.78}$$

where $\Delta\vec{E}_2\vec{E}_1$ is the subtraction of optical path between $\vec{E}_2$ and $\vec{E}_1$. If we take $\vec{E}_2$ as a reference beam, the milled depth will introduce a phase delay for $\vec{E}_1$ of

$$\Delta\rho = -L_{\Delta\vec{E}_2\vec{E}_1} \cdot k, \tag{4.79}$$

where k is the wavenumber.

The phase mask and fabrication depth should satisfy the following condition

$$phase - 2\pi = -L_{\Delta\vec{E}_2\vec{E}_1} \cdot k, \tag{4.80}$$

where *phase* means the desired phase mask. Since the phase mask is represented in $0 - 2\pi$ range as in Fig. 4.26, a 2π period is deducted from the origin phase to make it negative so as to equal to the phase delay on the right hand side.

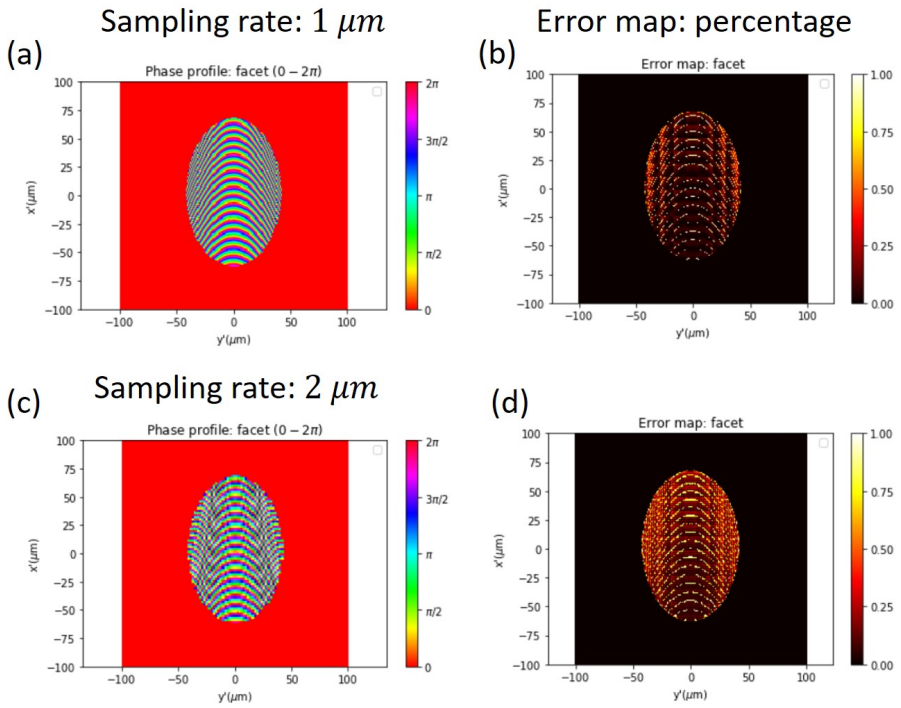

Figure 4.28: Discretized phase profile for the original phase mask in Fig. 4.26(d). (a) The phase profile for 1-μm resolution. I divide the whole phase map into a meshgrid with a lattice size of $1 \times 1 \mu m^2$. For each lattice, I take the averaged phase as the final phase profile. (b) The deviation error of the discretized phase profile in (a) and the original phase profile. (c) The phase profile for 2-μm resolution. (d) The deviation error of the discretized phase profile in (c) and the original phase profile.

We can calculate the fabricated depth as

$$d = (2\pi - phase) \cdot \frac{\cos \eta}{1 - \sin(2\eta - 90^\circ)} \cdot \frac{1}{k \cdot n_{NCF}}. \tag{4.81}$$

Fabrication

I will use focused ion beam (FIB) to mill different depth structure onto the fiber facet. The FIB machine I used is FEI Quanta 200 3D FIB from HKU Electron Microscope Unit (EMU).

Firstly, the depth profile of the phase mask in Fig. 4.26(d) is calculated from Eq. 4.81 and plotted in Fig. 4.30. The highest depth is 0.704 μm.

From the depth profile, I choose 6 depth steps to fabricate, namely, they are 150 nm, 250 nm, 350 nm, 450 nm, 550 nm, and 650 nm. For each pixel in the depth profile, they are rounded to the nearest step. Depth of 50 nm is eliminated, for it is beyond fabrication capacity. For each depth, its pattern is shown in Fig. 4.31.

The fabricated devices are shown in Fig. 4.32. I firstly milled the depth profile on a Silicon chip to test the validity. Then I milled the structure onto the NCF fiber facet.

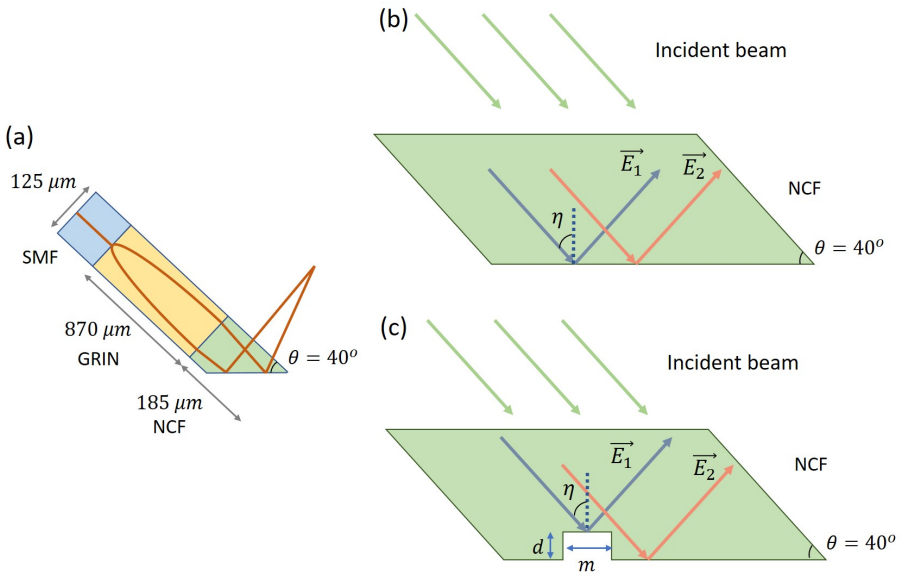

Figure 4.29: Schematic demonstration of introducing a phase mask through different structure. (a) Repositioning of the probe to let facet face downwards. (b) No structure is applied. The beam will go through a total internal reflection. There will be no phase shifting for each ray. (c) A depth is milled at one point. Compared with beam $\vec{E}_1$, beam $\vec{E}_2$ has a phase shift. d is the depth. m is the resolution, which is the size of one pixel of the whole structure. η is the incidence angle.

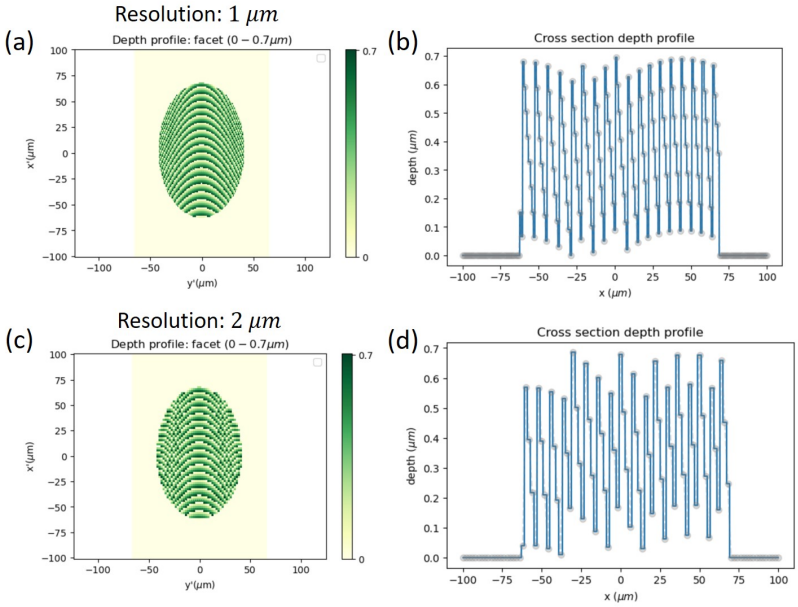

Figure 4.30: Depth profile of the phase profile in Fig. 4.26(d), whose working distance is 2000 μm. (a) Depth profile for resolution at 1 μm. (b) Cross section of depth profile in (a) at y=0. (c) Depth profile for resolution at 2 μm. (d) Cross section of depth profile in (c) at y=0.

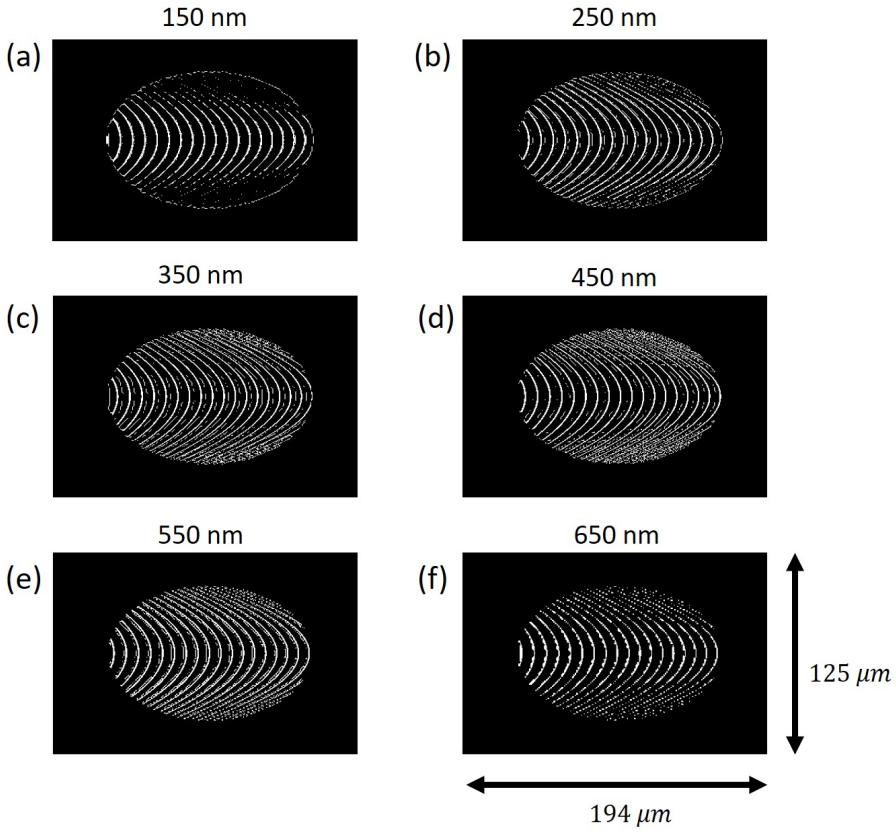

Figure 4.31: Depth profile for fabrication for phase profile in Fig. 4.26(d) with a working distance of 2000 μm. Black region will be untouched. White region will be milled at depth of 150 nm for pattern in (a), 250 nm for (b), 350 nm for (c), 450 nm for (d), 550 nm for (e), and 650 nm for (f).

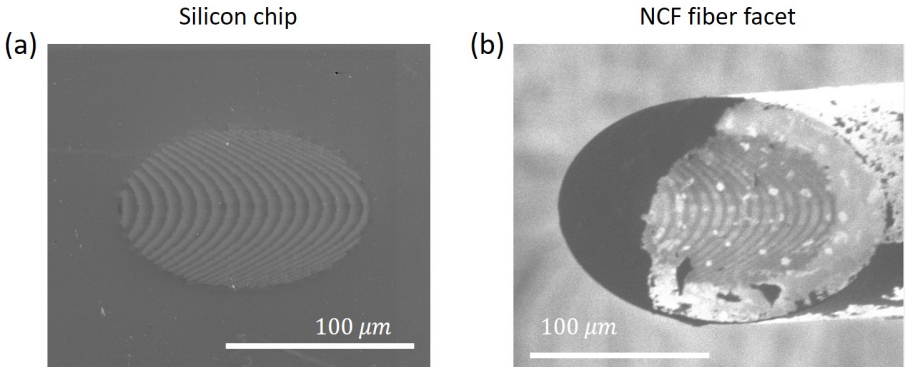

Figure 4.32: Fabricated device. (a) The depth profile is milled on a Silicon chip to test the validity. (2) The depth profile is milled on a NCF fiber facet.

Summary and discussion

All in all, this section presents the method to fabricate structures on a NCF fiber facet thus achieving an extended depth of focus. In the meta-probe, a GRIN fiber is utilized

to expand the beam and a NCF fiber is polished with an angle of 40° to facilitate total internal reflection. Due to the cylindrical shape of the fiber, the output beam will suffer from a astigmatism aberration. I calculated the phase profile we need to introduce on top of the facet to deliver a beam with an EDOF. The phase profile can be achieved through different depth, which could be milled through FIB. Sample fabricated devices are presented with a working distance of 2000 μm. Further testing will be conducted in the future.

However, there are some concerns about the geometrical optics solution. (1) In our phase profile, I select the resolution to be 1 μm, which is smaller than the wavelength 1.31 μm. The coupling effect may not be negligible in this circumstances, which may weaken the performance. (2) If we choose the resolution to be 2 μm, the total points may not be enough to represent the whole phase profile, which may also lead to a deteriorating performance. (3) As for total internal reflection, there is Goos-Hänchen effect [181]. The reflected beam will have a position shift, which means for every fabricated pixel depth, the reflected energy may not be 100% as the incident beam. It will cause a decreased efficiency. (4) The alignment precision during FIB fabrication is also a big challenge. Because the FIB milling will be conducted 6 times at different depth to create the complete profile, small misalignment between different fabrication will lead to a decreased performance.

4.6.3 Wave optics solution

This section presents the adjoint method to inversely design a structure on top of NCF fiber to create an EDOF for an enhanced biomedical imaging. The structure will be easy for FIB to fabricate. And the optimization method is based on full-wave FDTD simulation, which will take the coupling effect between different parts into consideration.

Problem definition

In physics, the problem can be phrased as how to design the structure on the NCF fiber facet to achieve an EDOF. In mathematics, the problem is actually a PDE problem in Eq. 4.17, which can be simplified as

$$\begin{aligned} \max_{\overleftrightarrow{\epsilon}} \quad & F(\mathbf{E}, \omega) \\ \textit{subject to} \quad & \nabla \cdot \overleftrightarrow{\epsilon}\mathbf{E} = \rho \\ & \nabla \cdot \overleftrightarrow{\mu}\mathbf{H} = 0 \\ & \nabla \times \mathbf{E} = -j\omega\overleftrightarrow{\mu}\mathbf{H} \\ & \nabla \times \mathbf{H} = \mathbf{J} + j\omega\overleftrightarrow{\epsilon}\mathbf{E}, \end{aligned} \tag{4.82}$$

where the tensor permittivity $\overleftrightarrow{\epsilon}$ represents material distribution. For the material in question, which will be silicon and silica in our case, the tensor permeability $\overleftrightarrow{\mu}$ is constant, thus being eliminated from the PDE problem. When solving the PDE problem

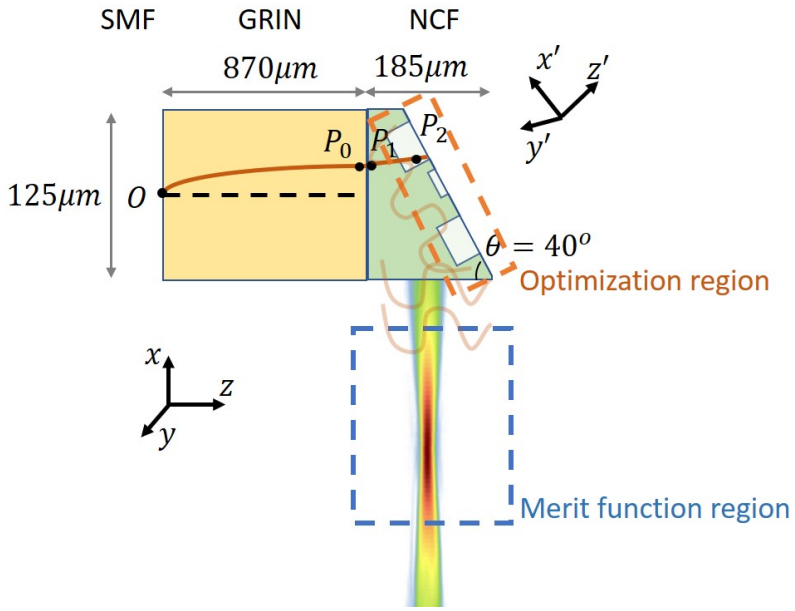

Figure 4.33: Schematic demonstration of inversely designing a structure on top of NCF fiber, which is based on full wave FDTD simulation. The incoming beam firstly interacts with the structure on the facet, which we are going to optimize. The reflected beam then will interfere with and transmit out of the fiber. A merit function region is chosen, within which we want to achieve an EDOF.

in Eq. 4.82, we try to find the optimal structure (material distribution) to maximize the predefined merit function.

The schematic demonstration of the fiber optimization is shown in Fig. 4.33. The optimization region, enclosed in the orange block, is the tensor permittivity distribution $\overleftrightarrow{\epsilon}$. The material distribution should accommodate the FIB fabrication. In other words, the structure should be similar to grating modality, which will be easier for top-to-down milling. The merit function region, enclosed in the blue block, represents the desired far field intensity distribution.

Optimization trial for a simplified case

To test the validity of the adjoint optimization method, a simplified case is demonstrated in Fig. 4.34.

It is a 2D silica structure with a refractive index of 1.45 at 1.31 μm. The incident source is a TE mode Gaussian source with a spectrum spanning from 1.19 μm to 1.41 μm and centered at 1.31 μm. We want to focus the beam at a point 14 μm above the the optimization region. The full wave FDTD simulation region will include the air area with a size of $33 \times 15 \mu m^2$. The FDTD simulation and adjoint optimization are conducted through Meep [49].

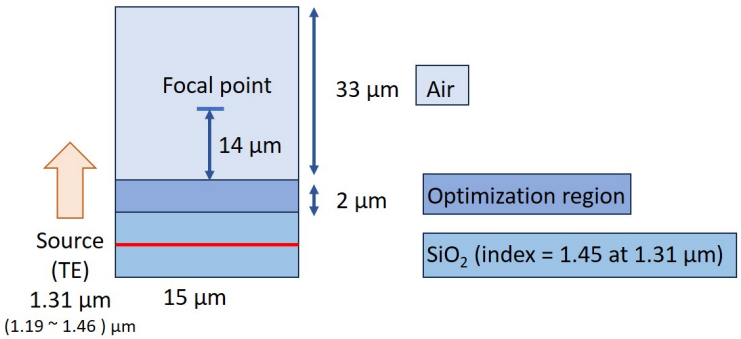

Figure 4.34: A simplified case for adjoint optimization. It is a 2D structure containing silica, whose refractive index is 1.45 at 1.31 μm. The red line represents the incident source, which is a TE source with a Gaussian specturm centered at 1.31 μm and spanning from 1.19 μm to 1.46 μm. The width of the structure is 15 μm. The optimization region is the block in the middle with a size of $15 \times 2\mu m^2$. We want to focus the incident beam at a point 14 μm higher above the optimization region.

We take n_f sampling points in the spectrum domain. The merit function is defined as

$$F = \sum_{i}^{n_f} ||\mathbf{E}(\mathbf{r}, \lambda_i)||^2, \tag{4.83}$$

where **r** is the predefined focal point.

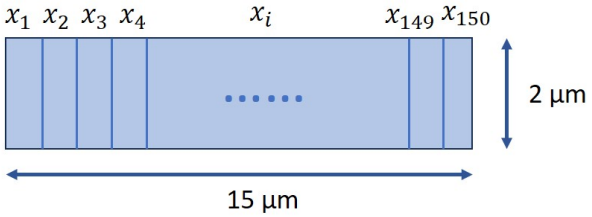

Figure 4.35: Grating modality for the optimization region. The 15-μm length has been equally divided into 150 parts. Each block is represented by a parameter x_i, whose value ranges from 0 (representing air) to 1 (representing silica). The whole optimization parameter is $\vec{x} = [x_1, x_2, x_3, ..., x_{150}]$.

To be compatible with fabrication, I choose grating as the modality in the optimization region. The resolution is set at 10 for 1-μm length. Since the length of the optimization region is 15 μm, there will be 150 blocks, shown in Fig. 4.35, with each being either air or silica. Mathematically, the optimization parameter can be expressed as

$$\vec{x} = [x_1, x_2, x_3, ..., x_{150}], \tag{4.84}$$

where x_i represents each block. The value of x_i ranges from 0 to 1. The value of 0 means air, while 1 represents silica. Since any other value between 0 and 1 is very hard to realize and fabricate, I will binarize the results after optimization.

Firstly, I will set the lower bounds for parameter $\vec{x}$ as

$$\underline{x} = [0, 0, 0, ..., 0], \tag{4.85}$$

and upper bounds as

$$\overline{x} = [1, 1, 1, ..., 1]. \tag{4.86}$$

For each parameter, the gradient $dF/d\vec{x}$ can be calculated through adjoint optimization, which is detailed in Ch. 4.2. Then the parameters will be updated as

$$\vec{x} := \vec{x} + \alpha \frac{dF}{d\vec{x}}, \tag{4.87}$$

where α is the step size. Because we want to maximize the merit function, we will add the gradient in the updating in Eq. 4.87.

After one optimization, I will apply a filter to the parameters to make them as binarized as possible for easier fabrication. The filter [182] can be expressed as

$$\vec{x} := \frac{\tanh(\beta \times \zeta) + \tanh(\beta \times (\vec{x} - \zeta))}{\tanh(\beta \times \zeta) + \tanh(\beta \times (1 - \zeta))}, \tag{4.88}$$

where β is the binarization factor initially set at 2, and ζ is the threshold with a value of 0.5. β will be multiplied by 2 after each iteration. The higher the value of β, the more binarized value the output will be. ζ is the dividing threshold, above which the value will be pushed to 1 and under which the value will be pushed to 0.

Then, we will go back to Eq. 4.87 to start a new round of optimization. β will be multiplied by 2 for 12 times. For each value of β, the optimization will run 6 times before increasing the β value. The total iteration loop will be 72 loops.

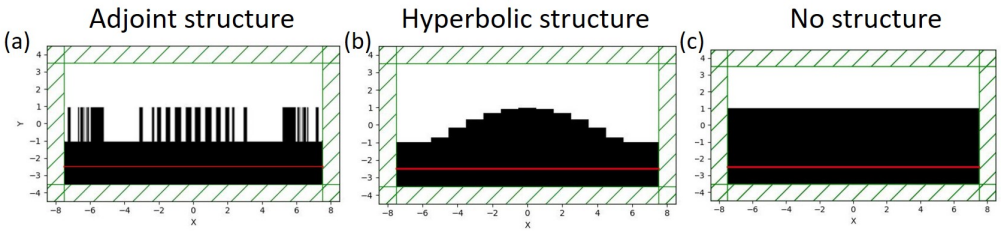

Figure 4.36: Structure comparison. (a) Optimized grating structure through adjoint methods. (b) Structure consisting of hyperbolic phase. (c) No structure. Red line represents the incident source.

When the optimization loop ends, I binarize the final parameter $\vec{x}$ with a threshold of 0.5, which in mathematical words is

$$\begin{aligned} &if\ x_i \geq 0.5,\ \ x_i := 1; \\ &if\ x_i < 0.5,\ \ x_i := 0. \end{aligned} \tag{4.89}$$

The finalized structure is shown in Fig. 4.36(a). For benchmark comparison, I also plot a structure with hyperbolic phase in Fig. 4.36(b) and an untouched structure in 4.36(c).

For the hyperbolic structure in Fig. 4.36(b), the 15-mum length is equally divided into 15 blocks. The height of each block is calculated as follows:

$$f \cdot n_{air} + d \cdot n_{SiO_2} = \sqrt{f^2 + r^2} \cdot n_{air} + d(r) \cdot n_{SiO_2}, \tag{4.90}$$

where f is the focal length of 14 μm we defined in Fig. 4.34, d is the height of the design region, r denotes the distance of a certain block from the center block, and $d(r)$ is the height for block at distance r. Therefore, the height for each block can be calculated as

$$d(r) = d + \left(f - \sqrt{f^2 + r^2}\right) \cdot \frac{n_{air}}{n_{SiO_2}}, \tag{4.91}$$

which is plotted in Fig. 4.36(b).

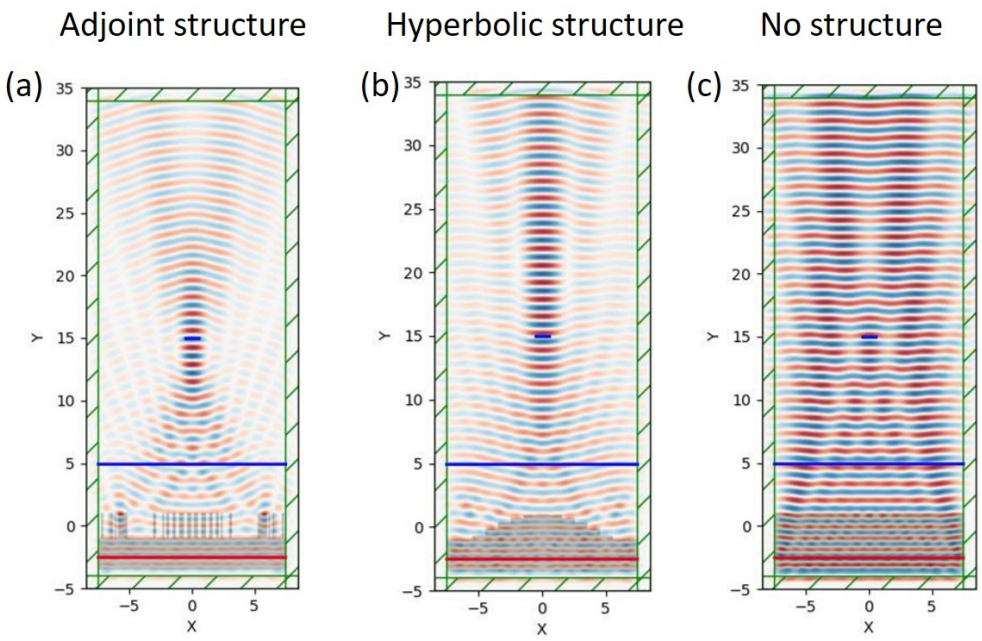

Figure 4.37: Electric field distribution for the three structures displayed in Fig. 4.36. Red line represents the incoming beam. Long blue line at 5 μm records the incoming flux. Short blue line at 15 μm records the flux at the focal point region.

For the three structure, an FDTD simulation is conducted to record the electric field response, which is shown in Fig. 4.37. It is obvious that adjoint structure and hyperbolic structure can focus the light at the desired point while no structure cannot perform the focusing function.

To quantitatively compare the results, I define a term as

$$focusing\ \ efficiency = \frac{focused\ \ power}{transmitted\ \ power} \tag{4.92}$$

The transmitted power is calculated as follows. After we run an FDTD simulation, we can have the electric field $\mathbf{E}(\vec{x},t)$ and magnetic field $\mathbf{H}(\vec{x},t)$ over space and time. For each point on the flux plane, we conduct the Fourier transform, which is

$$\begin{aligned}\mathbf{E}_\omega(\vec{x}) &= \frac{1}{\sqrt{2\pi}}\int e^{i\omega t}\mathbf{E}(\vec{x},t)dt \\ \mathbf{H}_\omega(\vec{x}) &= \frac{1}{\sqrt{2\pi}}\int e^{i\omega t}\mathbf{H}(\vec{x},t)dt.\end{aligned} \tag{4.93}$$

Then the power transmission is calculated through Poynting vector

$$P(\omega) = Re\Big(\hat{n}\cdot\int \mathbf{E}_\omega(\vec{x})\times\mathbf{H}_\omega(\vec{x})d^2\vec{x}\Big). \tag{4.94}$$

The focusing efficiency can be calculated as

$$focusing\ \ efficiency = \frac{P_{foc}(\omega)}{P_{tran}(\omega)}, \tag{4.95}$$

where $P_{foc}(\omega)$ denotes the power transmission spectrum on the focal plane, and $P_{tran}(\omega)$ denotes the power transmission spectrum on the transmission plane, which is the long blue line at 5 μm in Fig. 4.37.

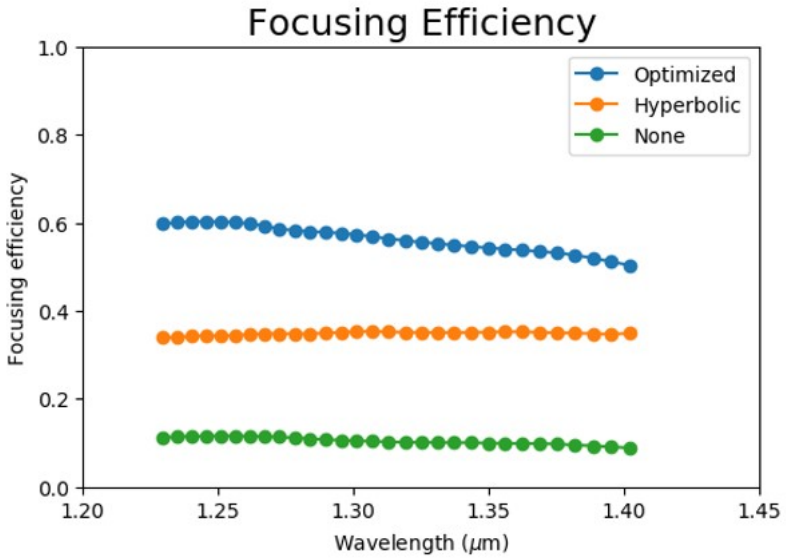

Figure 4.38: Focusing efficiency spectrum. Blue: adjoint optimized structure. Orange: hyperbolic structure. Green: no structure.

The focusing efficiency spectrum over 1.23 to 1.41 μm is shown in Fig. 4.38. The

hyperbolic structure has roughly 38% efficiency. The adjoint optimized structure has a much better performance, which has a 55% efficiency over the same spectrum. The results validate the advantage of adjoint optimization over traditional forward methods.

Optimization results for a large structure

This section demonstrates an optimized 3D grating structure on a silicon chip (with a refractive index of 3.47 at 1.7 μm) to achieve an EDOF. In the algorithm, we can readily switch the refractive index to silica. And the optimized structure can be fabricated on a flat NCF fiber, which will foster a forward view in an optical imaging setting.

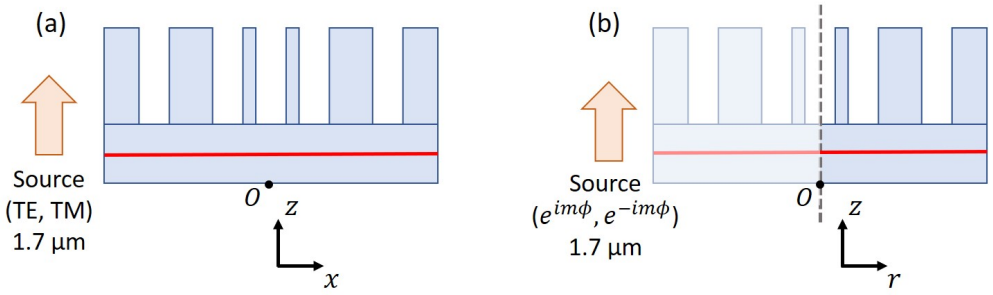

Figure 4.39: Full wave FDTD simulation in two coordinates. (a) Cartesian coordinate: one simulation has a TE source, while the other simulation has a TM source. (b) Cylindrical coordinate: one simulation has an LCP source with electric angular dependence of $e^{im\phi}$, while the other simulation has an RCP source with electric angular dependence of $e^{-im\phi}$.

I firstly run the adjoint optimization on a 2D structure, then expanding it to 3D structure through rotation symmetry. Such expansion needs some tactics shown in Fig. 4.39.

If we run it on a Cartesian coordinate [Fig. 4.39(a)], we have to run two simulations separately with a respective TE and TM source. The merit function should be altered to

$$F_{new} = F_{EDOF} - \delta||\mathbf{E}_{TE} - \mathbf{E}_{TM}||, \tag{4.96}$$

where F_{EDOF} denotes the old merit function to achieve an EDOF. And we want the electric field response of TE and TM source to be as small as possible, which is the second penalization term in Eq. 4.96. In the end, the optimized structure will be valid for any incoming polarized beam.

If we run it on a cylindrical coordinate [Fig. 4.39(b)], we also have to run two simulations separately with a respective LCP and RCP source. The LCP has an electric field angular dependence of $e^{im\phi}$, while LCP has an electric field angular dependence of $e^{-im\phi}$. The merit function should be changed to

$$F_{new} = F_{EDOF}(\mathbf{E}_{LCP} + \mathbf{E}_{RCP}). \tag{4.97}$$

A linear polarization is the sum of LCP and RCP. Therefore, the merit function in Eq. 4.97 will deliver an optimized solution for TE beam.

Here I choose to run the adjoint optimization in Cartesian coordinate with a TE source, which is shown in Fig. 4.40. The additional simulation with a TM source will be conducted in the future to deliver a more robust solution.

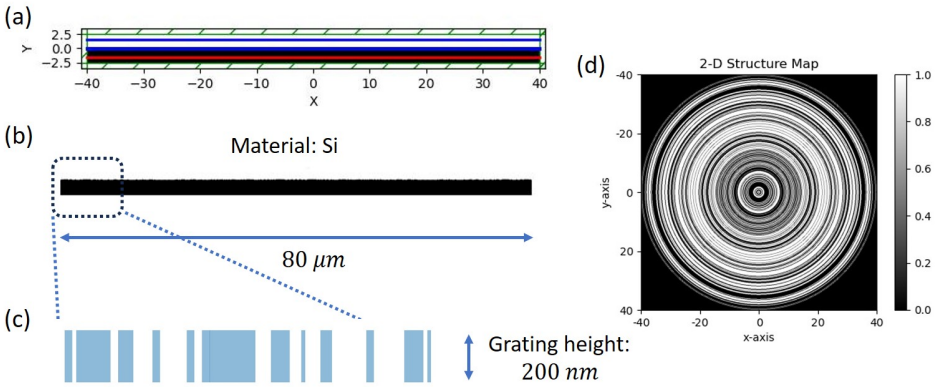

Figure 4.40: Optimization for a large structure. (a) Initial condition: design region has a length of 80 μm and a height of 200 nm. The incoming source is TE mode. (b) The optimized and binarized grating structure. The material is Si. (c) Zoomed in demonstration of the structure in (b). (d) The expanded structure through rotation symmetry from (b). Black region should be untouched, which white region shall be milled at 200-nm depth.

The optimization structure is demonstrated in Fig. 4.40(a). The design region has a length of 80 μm and a height of 200 nm. The incoming source is TE mode. After applying the adjoint methods, the optimized structure is shown in Fig. 4.40(b) with a zoomed-in figure in Fig. 4.40(c). Such 2D grating structure can be expanded to 3D through rotation symmetry with a pattern demonstrated in Fig. 4.40(d).

The merit function I want to maximize is

$$F = \sum_{i=1}^{20} ||\mathbf{E}_{A_i}||^2 - \sum_{i=1}^{10} ||\mathbf{E}_{S_{L_i}}||^2 - \sum_{i=1}^{10} ||\mathbf{E}_{S_{R_i}}||^2, \tag{4.98}$$

where $\mathbf{E}_{A_i}$, $\mathbf{E}_{S_{L_i}}$, and $\mathbf{E}_{S_{R_i}}$ denotes the electric field at points A_i, S_{L_i}, S_{R_i} respectively. The point A_i is in center region where we want the intensity to be maximized, while point S_{L_i} and S_{r_i} is in the side lobe region where the intensity is unwanted. So I will subtract the intensity in side lobe region in merit function. The points distribution is illustrated in Fig. 4.41(a). They have a coordinate as follows:

$$\begin{aligned} A_i &: (0, 500 + (i-1) \times 5), \ (i = 1, 2, ..., 20) \\ S_{L_i} &: (-10, 500 + (i-1) \times 10), \ (i = 1, 2, ..., 10) \\ S_{R_i} &: (10, 500 + (i-1) \times 10), \ (i = 1, 2, ..., 10) \end{aligned} \tag{4.99}$$

We conduct the same optimization loop as in last session Ch. 4.6.3. The optimized

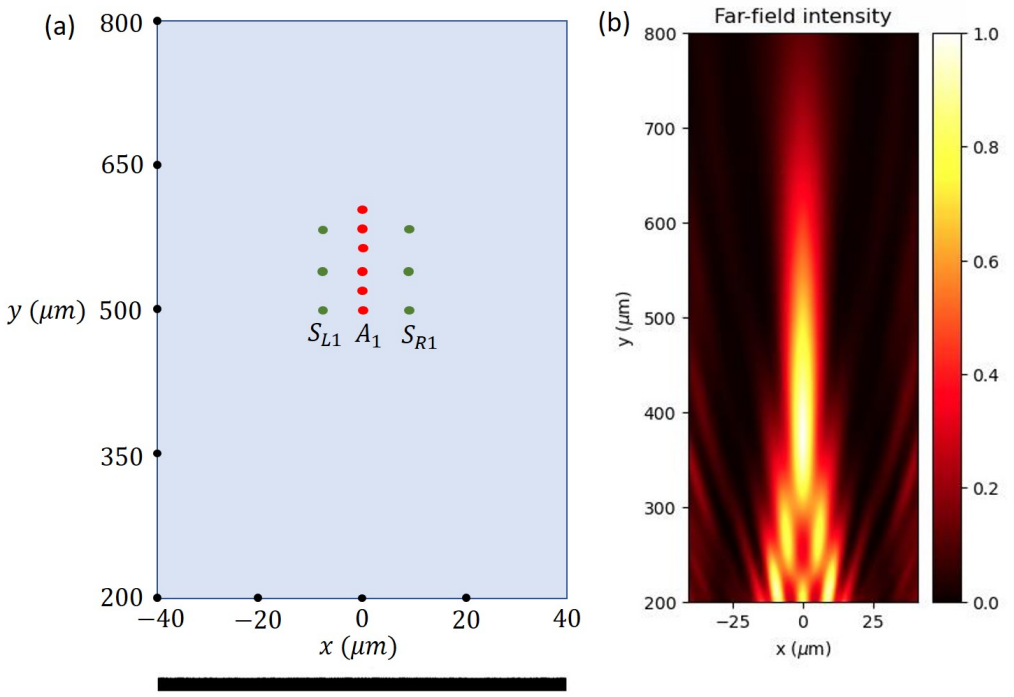

Figure 4.41: Far field intensity distribution. (a) Merit function sampling points distribution. Red points A_i is in center region, where the intensity will be maximized. Green points S_{L_i} and S_{R_i} is in the side lobe region, where the intensity will be minimized. The bottom structure is the optimized structure in Fig. 4.40(b). (b) Full wave FDTD simulation for the optimized structure. The far field intensity is normalized to 0-1 range.

grating structure is indicated in Fig. 4.40. Then I conduct the full wave FDTD simulation in the whole region to calculate the far field intensity, which is displayed in Fig. 4.41. We can notice an extended depth of focus from 350 μm to 700 μm, which validate our method on a large scale. In the next step, I will run the optimization with both TE and TM mode source so as to deliver a robust case when the structure is expanded to 3D.

Proposed optimization on a fiber with a side view

Previous optimization results are all on a plane, which can be applied to a flat fiber end face to achieve an enhance forward view in optical imaging. This section proposes the optimization method to achieve an EDOF on the side view as in Fig. 4.42.

The full wave FDTD simulation region shall cover the whole NCF fiber facet. Since the fiber is symmetric to the plane specified in Fig. 4.42, the simulation region can be halved. I will record the electric and magnetic field of the bottom plane of the simulation box in Fig. 4.42. An algorithm will be conducted to calculate the far field in those red points from the near field of the bottom plane, thus a merit function will be formed. (Here, I understand all six planes will contribute to the far field. I assume the bottom plane has the major contribution and the other five planes' contribution is negligible.)

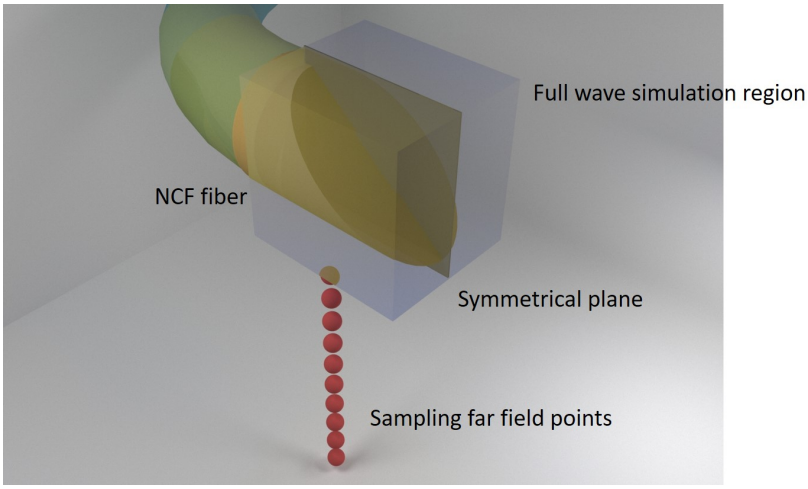

Figure 4.42: Proposed 3D optimization for a meta-fiber. The full wave FDTD simulation region is enclosed by the purple box. Red dots represent the far field points selection in the merit function. The NCF fiber is symmetrical to the plane specified in the figure, which can halve the simulation region.

Then, the adjoint optimization in Ch. 4.2 will be applied to compute the gradient upon each voxel. Next, the refractive index of the voxel will be updated towards either air or silica. And I will start a new optimization again. After multiple iterations, a final solution will be delivered in the end.

4.7 Proposed testing setup

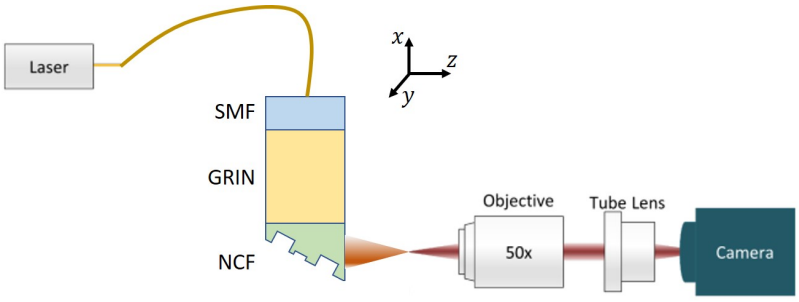

Figure 4.43: Proposed testing setup for the meta-fiber.

This section presents the experimental assessment of the beam characterization. The setup is shown in Fig. 4.43, which is inspired by literature [42].

I will optically pump the meta-fiber with a laser at a wavelength specified in previous optimization algorithm. The output beam profile at each axial points will be recorded by a camera paired with a lens system including a 50× objective and a tube lens. I will conduct the testing in the future.

4.8 Discussion

Here we utilized an adjoint optimization method to design an optical probe with an EDOF, which will potentially deliver an enhanced imaging result in clinical diagnosis.

The current endoscope OCT is too big. Our method can design a pattern directly on a fiber facet, which will significantly reduce the size. It will pave the way to scan small blood vessels. Still, lots of engineering problems will appear when we want to integrate an optical probe with electronic imaging devices. We have to consider about many factors, including the optical probe control, opto-electronic and electro-optic conversion, data receiving, processing, and storage, and even more. Although the engineering part including the fiber fabrication and system integration can be intimidating, the benefit of an enhanced biomedical imaging is crucial, which will largely improve the disease diagnosis ability.

4.9 Summary

All in all, I applied an adjoint optimization algorithm to inversely design a grating structure on a fiber facet to achieve an EDOF for an enhanced biomedical imaging. The equivalence of adjoint method and backpropagation in machine learning has been well explained in the theory part. I also developed an attenuation-compensation algorithm in OCT imaging, which has improved the imaging quality with a 8% increase in PIQE score. The adjoint optimization is conducted on both chip-modality and fiber-modality. For chip-modality, a phase profile has been obtained to achieve an EDOF with a central intensity 40% higher than its Bessel beam counterparts and a reduced side lobe. For fiber-modality, a grating structure on an oblique NCF fiber facet has been designed based on geometrical optics. The wave optics solution is still underway. I have applied the wave optics adjoint optimization to small structures, which reported an overall 10% increase in terms of focusing efficiency compared to traditional forward design methods. The optimization on a large plane structure also delivers a 350-μm EDOF. In the end, I propose an optimization loop for 3D NCF fiber and a testing setup to characterize the output beam, which will be performed in the future. The adjoint optimization has an advantage of converging a structure to satisfy the merit function. It requires less computation resources compared to the training data preparation in neural network architecture.

Chapter 5

Conclusion

5.1 Research contributions

Upon the completion of this thesis, huge efforts have been spent on the design and application of metasurfaces to improve "vision" in terms of signal "vision", holographic "vision", and biomedical "vision". The design technique ranges from forward calculation, machine learning prediction, to adjoint inverse design method. The three applications for "vision" improvement are just different representations of far-field radiation, which, through proper tuning, can achieve enhanced signal transmission, arbitrary holographic display, and augmented biomedical imaging.

For signal transmission, traditional optical communication methods usually adopt SMF, which only allows TE or TM mode beam to transmit. It will limit the communication speed with a maximum bottleneck of allowed orthogonal channels. I designed a metasurface to convert a plane wave into multiple LG modes and demultiplex them into different angles. Since LG mode has radial index and azimuthal index, the designed metasurface can expand the communication channels by one more dimension. Error analysis has also been conducted on this device to demonstrate a robust LG decomposition results over a broadband 400-nm wavelength range. It is envisaged that such technique will increase the current limit of optical communication speed to a new level. However, in order to achieve LG mode transmission in a fiber, it requires a micro-engineered fiber to receive and transmit LG modes. We will have to leverage the price of a tailored fiber and the benefit of an improved signal transmission speed.

For holographic display, current optical setting usually requires a Fourier transform lens to deliver a projection from a phase map. In order to achieve a lenless modality, the computation time to obtain a complex modulation mask is tremendous. Therefore, I developed a neural network to predict the complex mask upon the given image within only 1 second. The testing loss of this neural network is 4.57%, suggesting a good generalization. This technique has a huge potential to deliver a real-time and lensless holographic display, which has a promising application in augmented reality (AR) or virtual reality (VR). However, some engineering problems still exist. For example, we are using a metasurface with gold nano-structure. For a display application, it requires a structure

with at least 24 frame per second speed, e.g. SLM, liquid crystal. It is still a challenge for am SLM to deliver the same results as a static gold metasurface.

For biomedical imaging, cuurent imaging modality, e.g. OCT or photoacoustic imaging, suffers from limited working distance and output beam distortion. Therefore, I designed a meta-structure on a fiber to obtain an EDOF, which will increase the imaging resolution across depth. Here, the adjoint optimization method has been applied to design the fiber surface structure in an inverse way, which will outperform the traditional forward design method. The adjoint method has been tested on a small structure for validation, which demonstrates a more 10% improvement in focusing efficiency than traditional design. This technique will bring an enormous impact on clinical applications. Since different clinical imaging equipment have their own requirements, the adjoint inverse design method can converge to a probe design very quickly. However, lots of engineering problems will appear when we want to integrate an optical probe with electronic imaging devices. We have to consider about many factors, including the optical probe control, opto-electronic and electro-optic conversion, data receiving, processing, and storage, and more.

In sum, various metasurface design technique has been explored, from forward calculation, neural network, to adjoint inverse design methods. The finalized metasurface can be applied to signal transmission, holographic display, and biomedical imaging. It can be envisaged to improve the optical communication speed, deliver near-eye display, and enhance the clinical diagnosis. However, some engineering problems also exist for real life application as we discussed above.

Among the whole thesis, my research contribution includes idea conception, metasurface design and simulation, experimental setup, and data processing. The metasurface fabrication is accomplished by our collaborator Prof. Sun Xinkai, Dr. Yu Zejie, and Dr. Yu Yue in CUHK.

5.2 Future works

My current research can be further improved.

For signal transmission, one metasurface has been developed to increase the communication channels by one more dimension through LG modes. However, I only illustrated the beam profile when the metasurface was illuminated by the incoming beam. We can actually build a communication system with the current experimental setting as a transmitter. For the receiver part, I can choose suitable GRIN fibers to accommodate LG beam. And these fibers can be precisely positioned in space to receive the incoming LG signal. When these signals are detected, we can evaluate the loss through this transmission.

For the holographic display, I can only fabricate the metasurface as a static device, which cannot be tuned afterwards. The advantage of fast prediction through neural network cannot be manifested in this system. I am proposing we can apply the technique on an active device to achieve a real-time holographic display, which may have huge applications in the current augmented reality or virtual reality scenarios.

For biomedical imaging, here I only designed the grating structure on a fiber facet to achieve an EDOF. In the future, I will integrate the fiber with an imaging modality, be it OCT or photoacoustic imaging. Then I will utilize the enhanced imaging modality for in-vivo scanning.

Bibliography

[1] A. Salam. "Gauge Unification of the Four Fundamental Forces". In: *Physics and Contemporary Needs: Volume 2*. Springer, 1978, pp. 419–456.

[2] J. D. Jackson. *Classical electrodynamics*. American Association of Physics Teachers, 1999.

[3] E. Hecht. *Optics*. Pearson Education, 2012.

[4] F. T. Ulaby and U. Ravaioli. *Fundamentals of applied electromagnetics*. Vol. 7. Pearson Upper Saddle River, NJ, 2015.

[5] C. F. Gauss and C. F. Gauss. *Theoria attractionis corporum sphaeroidicorum ellipticorum homogeneorum: methodo nova tractata*. Springer, 1877.

[6] J. C. Maxwell. *A treatise on electricity and magnetism*. Vol. 1. Clarendon press, 1873.

[7] A. Garg. *Classical electromagnetism in a nutshell*. Vol. 13. Princeton University Press, 2012.

[8] D. P. Hampshire. "A derivation of Maxwell's equations using the Heaviside notation". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2134 (2018), p. 20170447.

[9] H. Hertz. *Electric waves: being researches on the propagation of electric action with finite velocity through space*. Dover Publications, 1893.

[10] D. B. Newell, E. Tiesinga, et al. "The international system of units (SI)". In: *NIST Special Publication* 330 (2019), pp. 1–138.

[11] J. W. Schmidt and M. R. Moldover. "Dielectric permittivity of eight gases measured with cross capacitors". In: *International Journal of Thermophysics* 24 (2003), pp. 375–403.

[12] M. Romer and I. B. Cohen. "Roemer and the First Determination of the Velocity of Light (1676)". In: *Isis* 31.2 (1940), pp. 327–379.

[13] M. Whybray. "Measuring the Speed of Light". In: (2011).

[14] S. Minardi, R. Harris, and L. Labadie. "Astrophotonics: astronomy and modern optics". In: *The Astronomy and Astrophysics Review* 29 (2021), pp. 1–81.

[15] M. G. Winkler and A. Van Helden. "Representing the heavens: Galileo and visual astronomy". In: *Isis* 83.2 (1992), pp. 195–217.

[16] M. Johns, P. McCarthy, K. Raybould, A. Bouchez, A. Farahani, J. Filgueira, G. Jacoby, S. Shectman, and M. Sheehan. "Giant magellan telescope: overview". In: *Ground-based and Airborne Telescopes IV* 8444 (2012), pp. 526–541.

[17] R. M. Wald. *General relativity*. University of Chicago press, 2010.

[18] G. D. Luker and K. E. Luker. "Optical imaging: current applications and future directions". In: *Journal of Nuclear Medicine* 49.1 (2008), pp. 1–4.

[19] J. Mertz. *Introduction to optical microscopy*. Cambridge University Press, 2019.

[20] R. M. Gagliardi and S. Karp. "Optical communications". In: *New York* (1976).

[21] B. L. Diffey. "Sources and measurement of ultraviolet radiation". In: *Methods* 28.1 (2002), pp. 4–13.

[22] L. Jin, Y.-W. Huang, Z. Jin, R. C. Devlin, Z. Dong, S. Mei, M. Jiang, W. T. Chen, Z. Wei, H. Liu, et al. "Dielectric multi-momentum meta-transformer in the visible". In: *Nature communications* 10.1 (2019), p. 4789.

[23] C. Jin, M. Afsharnia, R. Berlich, S. Fasold, C. Zou, D. Arslan, I. Staude, T. Pertsch, and F. Setzpfandt. "Dielectric metasurfaces for distance measurements and three-dimensional imaging". In: *Advanced Photonics* 1.3 (2019), p. 036001.

[24] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez. "Millimeter wave mobile communications for 5G cellular: It will work!" In: *IEEE access* 1 (2013), pp. 335–349.

[25] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, et al. "Deep learning for computer vision: A brief review". In: *Computational intelligence and neuroscience* 2018 (2018).

[26] S. D. Gupta, N. Ghosh, and A. Banerjee. *Wave optics: Basic concepts and contemporary trends*. CRC Press, 2015.

[27] A. F. Koenderink, A. Alu, and A. Polman. "Nanophotonics: Shrinking light-based technology". In: *Science* 348.6234 (2015), pp. 516–521.

[28] N. Zou. "Quantum entanglement and its application in quantum communication". In: *Journal of Physics: Conference Series*. Vol. 1827. 1. IOP Publishing. 2021, p. 012120.

[29] S. Daiss, S. Langenfeld, S. Welte, E. Distante, P. Thomas, L. Hartung, O. Morin, and G. Rempe. "A quantum-logic gate between distant quantum-network modules". In: *Science* 371.6529 (2021), pp. 614–617.

[30] C. Canalias and V. Pasiskevicius. "Mirrorless optical parametric oscillator". In: *Nature Photonics* 1.8 (2007), pp. 459–462.

[31] J. W. Lichtman and J.-A. Conchello. "Fluorescence microscopy". In: *Nature methods* 2.12 (2005), pp. 910–919.

[32] S. I. Bozhevolnyi, P. Genevet, and F. Ding. *Metasurfaces: Physics and Applications*. MDPI, 2018.

[33] N. Yu, P. Genevet, M. A. Kats, F. Aieta, J.-P. Tetienne, F. Capasso, and Z. Gaburro. "Light propagation with phase discontinuities: generalized laws of reflection and refraction". In: *science* 334.6054 (2011), pp. 333–337.

[34] J. Hu, S. Bandyopadhyay, Y.-h. Liu, and L.-y. Shao. "A review on metasurface: from principle to smart metadevices". In: *Frontiers in Physics* 8 (2021), p. 586087.

[35] A. L. Holsteen. "Actively Controlled Metasurfaces for Light Manipulation and Control". PhD thesis. Stanford University, 2017.

[36] M. Y. Shalaginov, S. An, F. Yang, P. Su, D. Lyzwa, A. M. Agarwal, H. Zhang, J. Hu, and T. Gu. "Single-element diffraction-limited fisheye metalens". In: *Nano Letters* 20.10 (2020), pp. 7429–7437.

[37] C. Chen, X. Ye, J. Sun, Y. Chen, C. Huang, X. Xiao, W. Song, S. Zhu, and T. Li. "Bifacial-metasurface-enabled pancake metalens with polarized space folding". In: *Optica* 9.12 (2022), pp. 1314–1322.

[38] L. Thibos, A. Bradley, D. Still, X. Zhang, and P. Howarth. "Theory and measurement of ocular chromatic aberration". In: *Vision research* 30.1 (1990), pp. 33–49.

[39] C. Novis. "Astigmatism and toric intraocular lenses". In: *Current opinion in ophthalmology* 11.1 (2000), pp. 47–50.

[40] M. A. Nanavaty, D. J. Spalton, and J. Marshall. "Effect of intraocular lens asphericity on vertical coma aberration". In: *Journal of Cataract & Refractive Surgery* 36.2 (2010), pp. 215–221.

[41] F. Balli, M. Sultan, S. K. Lami, and J. T. Hastings. "A hybrid achromatic metalens". In: *Nature communications* 11.1 (2020), p. 3892.

[42] H. Pahlevaninezhad, M. Khorasaninejad, Y.-W. Huang, Z. Shi, L. P. Hariri, D. C. Adams, V. Ding, A. Zhu, C.-W. Qiu, F. Capasso, et al. "Nano-optic endoscope for high-resolution optical coherence tomography in vivo". In: *Nature photonics* 12.9 (2018), pp. 540–547.

[43] Y. Bao, Y. Yu, H. Xu, C. Guo, J. Li, S. Sun, Z.-K. Zhou, C.-W. Qiu, and X.-H. Wang. "Full-colour nanoprint-hologram synchronous metasurface with arbitrary hue-saturation-brightness control". In: *Light: Science & Applications* 8.1 (2019), p. 95.

[44] A. C. Overvig, S. Shrestha, S. C. Malek, M. Lu, A. Stein, C. Zheng, and N. Yu. "Dielectric metasurfaces for complete and independent control of the optical amplitude and phase". In: *Light: Science & Applications* 8.1 (2019), p. 92.

[45] N. Mohammadi Estakhri, B. Edwards, and N. Engheta. "Inverse-designed metastructures that solve equations". In: *Science* 363.6433 (2019), pp. 1333–1338.

[46] E. Arbabi, A. Arbabi, S. M. Kamali, Y. Horie, M. Faraji-Dana, and A. Faraon. "MEMS-tunable dielectric metasurface lens". In: *Nature communications* 9.1 (2018), p. 812.

[47] S. C. Malek, H.-S. Ee, and R. Agarwal. "Strain multiplexed metasurface holograms on a stretchable substrate". In: *Nano letters* 17.6 (2017), pp. 3641–3645.

[48] F. Biancalana, A. Amann, A. V. Uskov, and E. P. O'reilly. "Dynamics of light propagation in spatiotemporal dielectric structures". In: *Physical Review E* 75.4 (2007), p. 046607.

[49] A. F. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J. D. Joannopoulos, and S. G. Johnson. "MEEP: A flexible free-software package for electromagnetic simulations by the FDTD method". In: *Computer Physics Communications* 181.3 (2010), pp. 687–702.

[50] K. Yee. "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media". In: *IEEE Transactions on antennas and propagation* 14.3 (1966), pp. 302–307.

[51] W. N. Plick and M. Krenn. "Physical meaning of the radial index of Laguerre-Gauss beams". In: *Phys. Rev. A* 92 (6 Dec. 2015), p. 063841. DOI: 10.1103/PhysRevA.92.063841. URL: https://link.aps.org/doi/10.1103/PhysRevA.92.063841.

[52] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman. "Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes". In: *Phys. Rev. A* 45 (11 June 1992), pp. 8185–8189. DOI: 10.1103/PhysRevA.45.8185. URL: https://link.aps.org/doi/10.1103/PhysRevA.45.8185.

[53] A. Faßbender, J. Babocký, P. Dvořák, V. Křápek, and S. Linden. "Invited Article: Direct phase mapping of broadband Laguerre-Gaussian metasurfaces". In: *APL Photonics* 3.11 (2018), p. 110803. DOI: 10.1063/1.5049368. eprint: https://doi.org/10.1063/1.5049368. URL: https://doi.org/10.1063/1.5049368.

[54] N. K. Fontaine, R. Ryf, H. Chen, D. T. Neilson, K. Kim, and J. Carpenter. "Laguerre-Gaussian mode sorter". In: *Nature Communications* 10.1 (2019), pp. 1–7.

[55] Y. Xiao, X. Tang, C. Wan, Y. Qin, H. Peng, C. Hu, and B. Qin. "Laguerre-Gaussian mode expansion for arbitrary optical fields using a subspace projection method". In: *Optics Letters* 44.7 (2019), pp. 1615–1618.

[56] Y. Wang, X. Fang, Z. Kuang, H. Wang, D. Wei, Y. Liang, Q. Wang, T. Xu, Y. Zhang, and M. Xiao. "On-chip generation of broadband high-order Laguerre-Gaussian modes in a metasurface". In: *Optics Letters* 42.13 (2017), pp. 2463–2466.

[57] G. Xie, H. Song, Z. Zhao, G. Milione, Y. Ren, C. Liu, R. Zhang, C. Bao, L. Li, Z. Wang, K. Pang, D. Starodubov, B. Lynn, M. Tur, and A. E. Willner. "Using a complex optical orbital-angular-momentum spectrum to measure object parameters". In: *Optics Letters* 42.21 (2017), pp. 4482–4485.

[58] M. P. Lavery, F. C. Speirits, S. M. Barnett, and M. J. Padgett. "Detection of a spinning object using light's orbital angular momentum". In: *Science* 341.6145 (2013), pp. 537–540.

[59] A. E. Willner, H. Huang, Y. Yan, Y. Ren, N. Ahmed, G. Xie, C. Bao, L. Li, Y. Cao, Z. Zhao, J. Wang, M. P. Lavery, M. Tur, S. Ramachandran, F. A. Molisch, N. Ashrafi, and S. Ashrafi. "Optical communications using orbital angular momentum beams". In: *Advances in Optics and Photonics* 7.1 (2015), pp. 66–106.

[60] H. Ren, G. Briere, X. Fang, P. Ni, R. Sawant, S. Héron, S. Chenot, S. Vézian, B. Damilano, V. Brändli, S. A. Maier, and P. Genevet. "Metasurface orbital angular momentum holography". In: *Nature Communications* 10.1 (2019), pp. 1–8.

[61] G. Gbur and R. K. Tyson. "Vortex beam propagation through atmospheric turbulence and topological charge conservation". In: *J. Opt. Soc. Am. A* 25.1 (2008), pp. 225–230.

[62] F. Yue, D. Wen, C. Zhang, B. D. Gerardot, W. Wang, S. Zhang, and X. Chen. "Multichannel polarization-controllable superpositions of orbital angular momentum states". In: *Advanced Materials* 29.15 (2017), p. 1603838.

[63] S. Li, X. Li, L. Zhang, G. Wang, L. Zhang, M. Liu, C. Zeng, L. Wang, Q. Sun, W. Zhao, and W. Zhang. "Efficient optical angular momentum manipulation for compact multiplexing and demultiplexing using a dielectric metasurface". In: *Advanced Optical Materials* 8.8 (2020), p. 1901666.

[64] E. Karimi, R. Boyd, P. De La Hoz, H. De Guise, J. Řeháček, Z. Hradil, A. Aiello, G. Leuchs, and L. L. Sánchez-Soto. "Radial quantum number of Laguerre-Gauss modes". In: *Physical Review A* 89.6 (2014), p. 063813.

[65] K. Sueda, G. Miyaji, N. Miyanaga, and M. Nakatsuka. "Laguerre-Gaussian beam generated with a multilevel spiral phase plate for high intensity laser pulses". In: *Optics Express* 12.15 (2004), pp. 3548–3553.

[66] H. Sztul and R. Alfano. "Double-slit interference with Laguerre-Gaussian beams". In: *Optics Letters* 31.7 (2006), pp. 999–1001.

[67] H. Luo, Z. Ren, W. Shu, and S. Wen. "Reversed propagation dynamics of Laguerre-Gaussian beams in left-handed materials". In: *Physical Review A* 77.2 (2008), p. 023812.

[68] S. Khonina, V. Kotlyar, R. Skidanov, V. Soifer, P. Laakkonen, and J. Turunen. "Gauss-Laguerre modes with different indices in prescribed diffraction orders of a diffractive phase element". In: *Optics Communications* 175.4-6 (2000), pp. 301–308.

[69] C. Wan, J. Chen, and Q. Zhan. "Tailoring optical orbital angular momentum spectrum with spiral complex field modulation". In: *Optics Express* 25.13 (2017), pp. 15108–15117.

[70] M. Rafayelyan and E. Brasselet. "Laguerre Gaussian modal q-plates". In: *Opt. Lett.* 42.10 (May 2017), pp. 1966–1969. DOI: 10.1364/OL.42.001966. URL: http://ol.osa.org/abstract.cfm?URI=ol-42-10-1966.

[71] M. Rafayelyan, T. Gertus, and E. Brasselet. "Laguerre-Gaussian quasi-modal q-plates from nanostructured glasses". In: *Applied Physics Letters* 110.26 (2017), p. 261108. DOI: 10.1063/1.4990954. eprint: https://doi.org/10.1063/1.4990954. URL: https://doi.org/10.1063/1.4990954.

[72] X. Ni, N. K. Emani, A. V. Kildishev, A. Boltasseva, and V. M. Shalaev. "Broadband light bending with plasmonic nanoantennas". In: *Science* 335.6067 (2012), pp. 427–427.

[73] X. Chen, L. Huang, H. Muhlenbernd, G. Li, B. Bai, Q. Tan, G. Jin, C.-W. Qiu, S. Zhang, and T. Zentgraf. "Dual-polarity plasmonic metalens for visible light". In: *Nature Communications* 3.1 (2012), pp. 1–6.

[74] D. Lin, P. Fan, E. Hasman, and M. L. Brongersma. "Dielectric gradient metasurface optical elements". In: *Science* 345.6194 (2014), pp. 298–302.

[75] B. Sang, Y. Ke, J. Wu, H. Luo, W. Shu, and S. Wen. "Generation of pure Laguerre-Gaussian vector beams on the higher-order Poincaré sphere by hollow Gaussian beams through dielectric metasurfaces". In: *Optics Communications* 439 (2019), pp. 27–33.

[76] Z. Bomzon, G. Biener, V. Kleiner, and E. Hasman. "Space-variant Pancharatnam-Berry phase optical elements with computer-generated subwavelength gratings". In: *Optics Letters* 27.13 (2002), pp. 1141–1143.

[77] M. Kang, T. Feng, H.-T. Wang, and J. Li. "Wave front engineering from an array of thin aperture antennas". In: *Optics Express* 20.14 (2012), pp. 15882–15890.

[78] L. Huang, X. Chen, H. Mühlenbernd, G. Li, B. Bai, Q. Tan, G. Jin, T. Zentgraf, and S. Zhang. "Dispersionless phase discontinuities for controlling light propagation". In: *Nano Letters* 12.11 (2012), pp. 5750–5755.

[79] G. Zheng, H. Mühlenbernd, M. Kenney, G. Li, T. Zentgraf, and S. Zhang. "Metasurface holograms reaching 80% efficiency". In: *Nature Nanotechnology* 10.4 (2015), pp. 308–312.

[80] H.-T. Chen, A. J. Taylor, and N. Yu. "A review of metasurfaces: physics and applications". In: *Reports on Progress in Physics* 79.7 (2016), p. 076401.

[81] G. Xie, Y. Ren, Y. Yan, H. Huang, N. Ahmed, L. Li, Z. Zhao, C. Bao, M. Tur, S. Ashrafi, and A. E. Willner. "Experimental demonstration of a 200-Gbit/s free-space optical link by multiplexing Laguerre-Gaussian beams with different radial indices". In: *Optics Letters* 41.15 (2016), pp. 3447–3450.

[82] M. Granata, C. Buy, R. Ward, and M. Barsuglia. "Higher-order Laguerre-Gauss mode generation and interferometry for gravitational wave detectors". In: *Physical review letters* 105.23 (2010), p. 231102.

[83] G. Ruffato, M. Massari, and F. Romanato. "Generation of high-order Laguerre–Gaussian modes by means of spiral phase plates". In: *Optics letters* 39.17 (2014), pp. 5094–5097.

[84] R. Borghi, F. Gori, and M. Santarsiero. "Optimization of Laguerre-Gauss truncated series". In: *Optics Communications* 125.4-6 (1996), pp. 197–203.

[85] M. M. P. Arnob, F. Zhao, J. Li, and W.-C. Shih. "EBL-based fabrication and different modeling approaches for nanoporous gold nanodisks". In: *ACS Photonics* 4.8 (2017), pp. 1870–1878.

[86] M. Feldman. *Nanolithography: the art of fabricating nanoelectronic and nanophotonic devices and systems*. Woodhead publishing, 2014.

[87] L. Novotny and B. Hecht. *Principles of nano-optics*. Cambridge university press, 2012.

[88] H. Mao. "first video LG42 LG11 6:4 with caption.mp4". In: (Aug. 2021). DOI: 10.6084/m9.figshare.14073833.v1. URL: https://opticapublishing.figshare.com/articles/media/first_video_LG42_LG11_6_4_with_caption_mp4/14073833.

[89] B. Picinbono. "On instantaneous amplitude and phase of signals". In: *IEEE Transactions on signal processing* 45.3 (1997), pp. 552–560.

[90] J. Lin, X.-C. Yuan, S. H. Tao, and R. E. Burge. "Collinear superposition of multiple helical beams generated by a single azimuthally modulated phase-only element". In: *Opt. Lett.* 30.24 (Dec. 2005), pp. 3266–3268. DOI: 10.1364/OL.30.003266. URL: http://ol.osa.org/abstract.cfm?URI=ol-30-24-3266.

[91] G. Molina-Terriza, J. P. Torres, and L. Torner. "Management of the angular momentum of light: preparation of photons in multidimensional vector states of angular momentum". In: *Physical Review Letters* 88.1 (2001), p. 013601.

[92] Y. Fazea and A. Amphawan. "40 Gbit/s MDM-WDM Laguerre-Gaussian Mode with Equalization for Multimode Fiber in Access Networks". In: *Journal of Optical Communications* 39.2 (2018), pp. 175–184. DOI: doi:10.1515/joc-2016-0138. URL: https://doi.org/10.1515/joc-2016-0138.

[93] J. Park, J.-H. Kang, S. J. Kim, X. Liu, and M. L. Brongersma. "Dynamic reflection phase and polarization control in metasurfaces". In: *Nano letters* 17.1 (2017), pp. 407–413.

[94] W. T. Chen, A. Y. Zhu, V. Sanjeev, M. Khorasaninejad, Z. Shi, E. Lee, and F. Capasso. "A broadband achromatic metalens for focusing and imaging in the visible". In: *Nature nanotechnology* 13.3 (2018), pp. 220–226.

[95] Z. Liu, D. Zhu, S. P. Rodrigues, K.-T. Lee, and W. Cai. "Generative model for the inverse design of metasurfaces". In: *Nano letters* 18.10 (2018), pp. 6570–6576.

[96] R. Lin, Y. Zhai, C. Xiong, and X. Li. "Inverse design of plasmonic metasurfaces by convolutional neural network". In: *Optics Letters* 45.6 (2020), pp. 1362–1365.

[97] W. Ma, Z. Liu, Z. A. Kudyshev, A. Boltasseva, W. Cai, and Y. Liu. "Deep learning for the design of photonic structures". In: *Nature Photonics* 15.2 (2021), pp. 77–90.

[98] J. Jiang and J. A. Fan. "Global optimization of dielectric metasurfaces using a physics-driven neural network". In: *Nano letters* 19.8 (2019), pp. 5366–5372.

[99] A. Sinha, J. Lee, S. Li, and G. Barbastathis. "Lensless computational imaging through deep learning". In: *Optica* 4.9 (2017), pp. 1117–1125.

[100] C. C. Nadell, B. Huang, J. M. Malof, and W. J. Padilla. "Deep learning for accelerated all-dielectric metasurface design". In: *Optics express* 27.20 (2019), pp. 27523–27535.

[101] X. Shi, T. Qiu, J. Wang, X. Zhao, and S. Qu. "Metasurface inverse design using machine learning approaches". In: *Journal of Physics D: Applied Physics* 53.27 (2020), p. 275105.

[102] S. An, C. Fowler, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, et al. "A deep learning approach for objective-driven all-dielectric metasurface design". In: *ACS Photonics* 6.12 (2019), pp. 3196–3207.

[103] Z. Zalevsky, D. Mendlovic, and R. G. Dorsch. "Gerchberg–Saxton algorithm applied in the fractional Fourier or the Fresnel domain". In: *Optics Letters* 21.12 (1996), pp. 842–844.

[104] J. W. Goodman. *Introduction to Fourier optics*. Roberts and Company publishers, 2005.

[105] H. Zhang, N. Collings, J. Chen, B. Crossland, D. Chu, and J. Xie. "Full parallax three-dimensional display with occlusion effect using computer generated hologram". In: *Optical Engineering* 50.7 (2011), pp. 074003–074003.

[106] X. Ding, F. Monticone, K. Zhang, L. Zhang, D. Gao, S. N. Burokur, A. De Lustrac, Q. Wu, C.-W. Qiu, and A. Alù. "Ultrathin Pancharatnam–Berry metasurface with maximal cross-polarization efficiency". In: *Advanced materials* 27.7 (2015), pp. 1195–1200.

[107] W. Zhao, H. Jiang, B. Liu, J. Song, Y. Jiang, C. Tang, and J. Li. "Dielectric Huygens' metasurface for high-efficiency hologram operating in transmission mode". In: *Scientific Reports* 6.1 (2016), p. 30613.

[108] W. Wan, J. Gao, and X. Yang. "Metasurface holograms for holographic imaging". In: *Advanced Optical Materials* 5.21 (2017), p. 1700541.

[109] X. Ni, A. V. Kildishev, and V. M. Shalaev. "Metasurface holograms for visible light". In: *Nature communications* 4.1 (2013), p. 2807.

[110] O. Stephen, U. J. Maduh, S. Ibrokhimov, K. L. Hui, A. A. Al-Absi, and M. Sain. "A multiple-loss dual-output convolutional neural network for fashion class classification". In: *2019 21st International Conference on Advanced Communication Technology (ICACT)*. IEEE. 2019, pp. 408–412.

[111] D. Gabor. "A new microscopic principle." In: (1948).

[112] D. Pi, J. Liu, and Y. Wang. "Review of computer-generated hologram algorithms for color dynamic holographic three-dimensional display". In: *Light: Science & Applications* 11.1 (2022), p. 231.

[113] T.-C. Poon and J.-P. Liu. *Introduction to modern digital holography: with MATLAB*. Cambridge University Press, 2014.

[114] J. C. Dainty. *Laser speckle and related phenomena*. Vol. 9. Springer science & business Media, 2013.

[115] J. I. Trisnadi. "Speckle contrast reduction in laser projection displays". In: *Projection displays VIII*. Vol. 4657. SPIE. 2002, pp. 131–137.

[116] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, et al. "U-Net: deep learning for cell counting, detection, and morphometry". In: *Nature methods* 16.1 (2019), pp. 67–70.

[117] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[118] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241.

[119] S. Ioffe and C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *International conference on machine learning*. pmlr. 2015, pp. 448–456.

[120] S. Mehta, C. Paunwala, and B. Vaidya. "CNN based traffic sign classification using Adam optimizer". In: *2019 international conference on intelligent computing and control systems (ICCS)*. IEEE. 2019, pp. 1293–1298.

[121] J. Yamanaka, S. Kuwashima, and T. Kurita. "Fast and accurate image super resolution by deep CNN with skip connection and network in network". In: *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II 24*. Springer. 2017, pp. 217–225.

[122] Y. Fan, Y. Xu, M. Qiu, W. Jin, L. Zhang, E. Y. Lam, D. P. Tsai, and D. Lei. "Phase-controlled metasurface design via optimized genetic algorithm". In: *Nanophotonics* 9.12 (2020), pp. 3931–3939.

[123] M. Zhou, D. Liu, S. W. Belling, H. Cheng, M. A. Kats, S. Fan, M. L. Povinelli, and Z. Yu. "Inverse design of metasurfaces based on coupled-mode theory and adjoint optimization". In: *ACS Photonics* 8.8 (2021), pp. 2265–2273.

[124] F. Chollet. *Deep learning with Python*. Simon and Schuster, 2021.

[125] A. B. E. Attia, G. Balasundaram, M. Moothanchery, U. Dinish, R. Bi, V. Ntziachristos, and M. Olivo. "A review of clinical photoacoustic imaging: Current and future trends". In: *Photoacoustics* 16 (2019), p. 100144.

[126] D. A. Long. "Raman spectroscopy". In: *New York* 1 (1977).

[127] L. Schermelleh, A. Ferrand, T. Huser, C. Eggeling, M. Sauer, O. Biehlmaier, and G. P. Drummen. "Super-resolution microscopy demystified". In: *Nature cell biology* 21.1 (2019), pp. 72–84.

[128] J. Zhao, Y. Winetraub, L. Du, A. Van Vleck, K. Ichimura, C. Huang, S. Z. Aasi, K. Y. Sarin, and A. de la Zerda. "Flexible method for generating needle-shaped beams and its application in optical coherence tomography". In: *Optica* 9.8 (2022), pp. 859–867.

[129] M. Siddiqui, A. S. Nam, S. Tozburun, N. Lippok, C. Blatter, and B. J. Vakoc. "High-speed optical coherence tomography by circular interferometric ranging". In: *Nature photonics* 12.2 (2018), pp. 111–116.

[130] R. Cao, J. Zhao, L. Li, L. Du, Y. Zhang, Y. Luo, L. Jiang, S. Davis, Q. Zhou, A. de la Zerda, et al. "Optical-resolution photoacoustic microscopy with a needle-shaped beam". In: *Nature Photonics* 17.1 (2023), pp. 89–95.

[131] K. S. Bronk, K. L. Michael, P. Pantano, and D. R. Walt. "Combined imaging and chemical sensing using a single optical imaging fiber". In: *Analytical chemistry* 67.17 (1995), pp. 2750–2757.

[132] F. He, J. Yu, Y. Tan, W. Chu, C. Zhou, Y. Cheng, and K. Sugioka. "Tailoring femtosecond 1.5-μm Bessel beams for manufacturing high-aspect-ratio through-silicon vias". In: *Scientific reports* 7.1 (2017), p. 40785.

[133] E. Bayati, R. Pestourie, S. Colburn, Z. Lin, S. G. Johnson, and A. Majumdar. "Inverse designed metalenses with extended depth of focus". In: *ACS photonics* 7.4 (2020), pp. 873–878.

[134] G. Di Domenico, G. Ruocco, C. Colosi, E. DelRe, and G. Antonacci. "Cancellation of Bessel beam side lobes for high-contrast light sheet microscopy". In: *Scientific reports* 8.1 (2018), p. 17178.

[135] L. Huang, J. Whitehead, S. Colburn, and A. Majumdar. "Design and analysis of extended depth of focus metalenses for achromatic computational imaging". In: *Photonics Research* 8.10 (2020), pp. 1613–1623.

[136] M. Mansouree, A. McClung, S. Samudrala, and A. Arbabi. "Large-scale parametrized metasurface design using adjoint optimization". In: *ACS Photonics* 8.2 (2021), pp. 455–463.

[137] Y. Mao, S. Chang, S. Sherif, and C. Flueraru. "Graded-index fiber lens proposed for ultrasmall probes used in biomedical imaging". In: *Applied optics* 46.23 (2007), pp. 5887–5894.

[138] D. Kim, J. Kim, Y. I. Park, N. Lee, and T. Hyeon. "Recent development of inorganic nanoparticles for biomedical imaging". In: *ACS central science* 4.3 (2018), pp. 324–336.

[139] R. M. Errico. "What is an adjoint model?" In: *Bulletin of the American Meteorological Society* 78.11 (1997), pp. 2577–2592.

[140] Y. Cao, S. Li, L. Petzold, and R. Serban. "Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution". In: *SIAM journal on scientific computing* 24.3 (2003), pp. 1076–1089.

[141] S.-i. Amari. "Backpropagation and stochastic gradient descent method". In: *Neurocomputing* 5.4-5 (1993), pp. 185–196.

[142] C. Olah. "Calculus on computational graphs: Backpropagation". In: (2015).

[143] R. Hecht-Nielsen. "Theory of the backpropagation neural network". In: *Neural networks for perception*. Elsevier, 1992, pp. 65–93.

[144] O. D. Miller. *Photonic design: From fundamental solar cell physics to computational inverse design*. University of California, Berkeley, 2012.

[145] Y. Zhao, W. J. Eldridge, J. R. Maher, S. Kim, M. Crose, M. Ibrahim, H. Levinson, and A. Wax. "Dual-axis optical coherence tomography for deep tissue imaging". In: *Optics letters* 42.12 (2017), pp. 2302–2305.

[146] M. T. Leite, L. M. Zangwill, R. N. Weinreb, H. L. Rao, L. M. Alencar, P. A. Sample, and F. A. Medeiros. "Effect of disease severity on the performance of Cirrus spectral-domain OCT for glaucoma diagnosis". In: *Investigative ophthalmology & visual science* 51.8 (2010), pp. 4104–4109.

[147] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, et al. "Optical coherence tomography". In: *science* 254.5035 (1991), pp. 1178–1181.

[148] G. J. Tearney, M. E. Brezinski, B. E. Bouma, S. A. Boppart, C. Pitris, J. F. Southern, and J. G. Fujimoto. "In vivo endoscopic optical biopsy with optical coherence tomography". In: *Science* 276.5321 (1997), pp. 2037–2039.

[149] A. Badon, D. Li, G. Lerosey, A. C. Boccara, M. Fink, and A. Aubry. "Smart optical coherence tomography for ultra-deep imaging through highly scattering media". In: *Science advances* 2.11 (2016), e1600370.

[150] N. M. Israelsen, C. R. Petersen, A. Barh, D. Jain, M. Jensen, G. Hannesschläger, P. Tidemand-Lichtenberg, C. Pedersen, A. Podoleanu, and O. Bang. "Real-time high-resolution mid-infrared optical coherence tomography". In: *Light: Science & Applications* 8.1 (2019), p. 11.

[151] W. Drexler and J. G. Fujimoto. *Optical coherence tomography: technology and applications*. Springer Science & Business Media, 2008.

[152] Y. Shimada, H. Nakagawa, A. Sadr, I. Wada, M. Nakajima, T. Nikaido, M. Otsuki, J. Tagami, and Y. Sumi. "Noninvasive cross-sectional imaging of proximal caries using swept-source optical coherence tomography (SS-OCT) in vivo". In: *Journal of biophotonics* 7.7 (2014), pp. 506–513.

[153] U. Baran, W. Qin, X. Qi, G. Kalkan, and R. K. Wang. "OCT-based label-free in vivo lymphangiography within human skin and areola". In: *Scientific reports* 6.1 (2016), p. 21122.

[154] N. Uribe-Patarroyo and B. E. Bouma. "Rotational distortion correction in endoscopic optical coherence tomography based on speckle decorrelation". In: *Optics letters* 40.23 (2015), pp. 5518–5521.

[155] S. Ortiz, D. Siedlecki, I. Grulkowski, L. Remon, D. Pascual, M. Wojtkowski, and S. Marcos. "Optical distortion correction in optical coherence tomography for quantitative ocular anterior segment by three-dimensional imaging". In: *Optics express* 18.3 (2010), pp. 2782–2796.

[156] S. Yoon, M. Kim, M. Jang, Y. Choi, W. Choi, S. Kang, and W. Choi. "Deep optical imaging within complex scattering media". In: *Nature Reviews Physics* 2.3 (2020), pp. 141–158.

[157] A. Golabchi, J. Faust, F. Golabchi, D. Brooks, A. Gouldstone, and C. DiMarzio. "Refractive errors and corrections for OCT images in an inflated lung phantom". In: *Biomedical optics express* 3.5 (2012), pp. 1101–1109.

[158] D. Levitz, L. Thrane, M. H. Frosz, P. E. Andersen, C. B. Andersen, J. Valanciunaite, J. Swartling, S. Andersson-Engels, and P. R. Hansen. "Determination of optical scattering properties of highly-scattering media in optical coherence tomography images". In: *Optics express* 12.2 (2004), pp. 249–259.

[159] K. A. Vermeer, J. Mo, J. J. Weda, H. G. Lemij, and J. F. de Boer. "Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography". In: *Biomedical optics express* 5.1 (2014), pp. 322–337.

[160] K. Minami, Y. Kataoka, J. Matsunaga, S. Ohtani, M. Honbou, and K. Miyata. "Ray-tracing intraocular lens power calculation using anterior segment optical coherence tomography measurements". In: *Journal of Cataract & Refractive Surgery* 38.10 (2012), pp. 1758–1763.

[161] A. de Castro, S. Barbero, S. Ortiz, and S. Marcos. "Accuracy of the reconstruction of the crystalline lens gradient index with optimization methods from Ray Tracing and Optical Coherence Tomography data". In: *Optics express* 19.20 (2011), pp. 19265–19279.

[162] L. Thrane, H. T. Yura, and P. E. Andersen. "Analysis of optical coherence tomography systems based on the extended Huygens–Fresnel principle". In: *JOSA A* 17.3 (2000), pp. 484–490.

[163] K. C. Zhou, R. Qian, S. Degan, S. Farsiu, and J. A. Izatt. "Optical coherence refraction tomography". In: *Nature photonics* 13.11 (2019), pp. 794–802.

[164] D. J. Smithies, T. Lindmo, Z. Chen, J. S. Nelson, and T. E. Milner. "Signal attenuation and localization in optical coherence tomography studied by Monte Carlo simulation". In: *Physics in Medicine & Biology* 43.10 (1998), p. 3025.

[165] D. J. Faber, F. J. Van Der Meer, M. C. Aalders, and T. G. van Leeuwen. "Quantitative measurement of attenuation coefficients of weakly scattering media using optical coherence tomography". In: *Optics express* 12.19 (2004), pp. 4353–4365.

[166] A. Mariampillai, M. K. Leung, M. Jarvi, B. A. Standish, K. Lee, B. C. Wilson, A. Vitkin, and V. X. Yang. "Optimized speckle variance OCT imaging of microvasculature". In: *Optics letters* 35.8 (2010), pp. 1257–1259.

[167] V. Y. Zaitsev, L. Matveev, A. Matveyev, A. Sovetsky, D. Shabanov, S. Ksenofontov, G. Gelikonov, O. Baum, A. Omelchenko, A. Yuzhakov, et al. "Optimization of phase-resolved optical coherence elastography for highly-sensitive monitoring of slow-rate strains". In: *Laser Physics Letters* 16.6 (2019), p. 065601.

[168] M. Casper, H. Schulz-Hildebrandt, M. Evers, R. Birngruber, D. Manstein, and G. Hüttmann. "Optimization-based vessel segmentation pipeline for robust quantification of capillary networks in skin with optical coherence tomography angiography". In: *Journal of Biomedical Optics* 24.4 (2019), pp. 046005–046005.

[169] M. J. Ju, S. J. Lee, E. J. Min, Y. Kim, H. Y. Kim, and B. H. Lee. "Evaluating and identifying pearls and their nuclei by using optical coherence tomography". In: *Optics express* 18.13 (2010), pp. 13468–13477.

[170] Y. Zhou, T. Liu, Y. Shi, Z. Chen, J. Mao, and W. Zhou. "Automated internal classification of beadless Chinese ZhuJi freshwater pearls based on optical coherence tomography images". In: *Scientific Reports* 6.1 (2016), p. 33819.

[171] M. J. Ju, S. J. Lee, Y. Kim, J. G. Shin, H. Y. Kim, Y. Lim, Y. Yasuno, and B. H. Lee. "Multimodal analysis of pearls and pearl treatments by using optical coherence tomography and fluorescence spectroscopy". In: *Optics express* 19.7 (2011), pp. 6420–6432.

[172] A. Y.-M. Lin, P.-Y. Chen, and M. A. Meyers. "The growth of nacre in the abalone shell". In: *Acta Biomaterialia* 4.1 (2008), pp. 131–138.

[173] J. Kang, P. Feng, X. Wei, E. Y. Lam, K. K. Tsia, and K. K. Wong. "102-nm, 44.5-MHz inertial-free swept source by mode-locked fiber laser and time stretch technique for optical coherence tomography". In: *Optics express* 26.4 (2018), pp. 4370–4381.

[174] R. Huber, M. Wojtkowski, and J. Fujimoto. "Fourier Domain Mode Locking (FDML): A new laser operating regime and applications for optical coherence tomography". In: *Optics express* 14.8 (2006), pp. 3225–3237.

[175] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani. "Blind image quality evaluation using perception based features". In: *2015 twenty first national conference on communications (NCC)*. IEEE. 2015, pp. 1–6.

[176] D. Micieli, D. Di Martino, M. Musa, L. Gori, A. Kaestner, A. Bravin, A. Mittone, R. Navone, and G. Gorini. "Characterizing pearls structures using X-ray phase-contrast and neutron imaging: a pilot study". In: *Scientific Reports* 8.1 (2018), p. 12118.

[177] D. Junqing and L. Qinghui. "Research on nondestructive measurement of sea pearls using optical coherence tomography technique". In: *Infrared and Laser Engineering* 47.4 (2018), pp. 417004–0417004.

[178] N. Xie, M. D. Carson, J. E. Fröch, A. Majumdar, E. J. Seibel, and K. F. Böhringer. "Large field-of-view short-wave infrared metalens for scanning fiber endoscopy". In: *Journal of Biomedical Optics* 28.9 (2023), pp. 094802–094802.

[179] J. Kang, R. Zhu, Y. Sun, J. Li, and K. K. Wong. "Pencil-beam scanning catheter for intracoronary optical coherence tomography". In: *Opto-Electronic Advances* 5.3 (2022), pp. 200050–1.

[180] B. E. Saleh and M. C. Teich. *Fundamentals of photonics*. john Wiley & sons, 2019.

[181] K. Chiu and J. Quinn. "On the Goos-Hänchen effect: a simple example of a time delay scattering process". In: *American Journal of Physics* 40.12 (1972), pp. 1847–1851.

[182] F. Wang, B. S. Lazarov, and O. Sigmund. "On projection methods, convergence and robust formulations in topology optimization". In: *Structural and multidisciplinary optimization* 43 (2011), pp. 767–784.

List of Abbreviations

SI	**S**ystème **I**nternational: international system of units
FEM	**F**inite **E**lement **M**ethod
FDTD	**F**inite-**D**ifference **T**ime-**D**omain
SEM	**S**canning **E**lectron **M**icroscope
FOV	**F**ield **O**f **V**iew
OCT	**O**ptical **C**oherence **T**omography
3D	**T**hree **D**imensional
MEMS	**M**icro**E**lectro**M**echanical **S**ystem
LG	**L**aguerre **G**aussian
OAM	**O**rbital **A**ngular **M**omentum
LCP	**L**eft **C**ircular **P**olarized
RCP	**R**ight **C**ircular **P**olarized
CGH	**C**omputer **G**enerated **H**ologram
FIB	**F**ocused **I**on **B**eam
ITO	**I**ndium **T**in **O**xide
EM	**E**lectro-**M**agnetic
SLM	**S**patial **L**ight **M**odulator
CGH	**C**omputer **G**enerated **H**ologram
NO	**N**eural **O**ptimizer
ReLU	**Re**ctified **L**inear **U**nit
BN	**B**atch **N**ormalization
MAE	**M**ean **A**bsolute **E**rror
CMT	**C**oupled-**M**ode **T**heory
PBS	**P**olarization **B**eam **S**plitter
CCD	**C**harge **C**oupled **D**evice
EBL	**E**lectron **B**eam **L**ithography
LFW	**L**abelled **F**ace in the **W**ild
EDOF	**E**xtended **D**epth **o**f **F**ocus
PA	**P**hoto**a**coustic
DOE	**D**iffractive **O**ptical **E**lement
DOF	**D**epth **o**f **F**ocus
GRIN	**Gr**aded **In**dex
SMF	**S**ingle **M**ode **F**iber
MRI	**M**agnetic **R**esonance **I**maging
CT	**C**omputed **T**omography

PDE	**P**artial **D**ifferential **E**quation
NCF	**N**on-**C**ore **F**iber
ODE	**O**rdinary **D**ifferential **E**quation
EMU	**E**lectron **M**icroscope **U**nit
PML	**P**erfectly **M**atched **L**ayer
CUHK	**C**hinese **U**niversity of **H**ong **K**ong

List of Symbols

Global notations

$\mathbf{E}$	electric field	—
$\mathbf{B}$	magnetic field	—
ϵ_0	permittivity in vacuum (8.85×10^{-12})	$s^2 \cdot C^2/m^3 \cdot kg$
μ_0	permeability in vacuum ($4\pi \times 10^{-7}$)	$m \cdot kg/C^2$
k	wavenumber	—
ω	angular frequency	rad

Chapter 1

ρ	electric charge density	—

Chapter 2

α	orientation angle of the block	—
L	left circular polarization	—
Γ	rotation matrix	—
Ψ	output field	—
ϕ_o	phase modulation from ordinary axis	—
ϕ_e	phase modulation from extraordinary axis	—
A_o	amplitude modulation from ordinary axis	—
A_e	amplitude modulation from extraordinary axis	—
d	height of the proposed Au block	—
A_{out}	output amplitude	—
ϕ_{out}	output phase	—
p	radial index	—
l	azimuthal index	—
ω_0	beam waist	—
r, ϕ	cylindrical coordinates	—
σ_s	LG mode rario	—
θ	deflection angle	—
$\vec{g}_s$	deflection direction	—
$\vec{\rho}$	pixel vector	—

U	complex field	—
C_p^l	LG mode decomposition coefficient	—
η_N	desired quasi-amplitude conversion	—
η_P	desired actual amplitude conversion	—
w_y	length of the nano-antenna along y-axis	nm
α_0	initial orientation angle	—
η_{ac}	actual amplitude conversion	—
ϕ_{ac}	actual phase conversion	—
ξ	error between actual and desired complex conversion	—
ξ_1	pessimistic error	—
ξ_2	neutral error	—
ξ_3	optimistic error	—
Z	vector of the length and width of Au block	—
di	dimension of the variable	—
Σ	covariance matrix of the Gaussian distribution	—
μ	mean of the length and width	—
ϵ	error map	—
p	probability density function distribution	—
σ_{pl}	complex coefficient of LG mode	—
γ	efficiency	—

Chapter 3

$\vec{r}$	original point coordinate	—
$\vec{r}_1$	target point coordinate	—
G_0	Green function	—
H	hologprahic image	—
M	complex modulation of a metasurface	—
R	distance between metasurface and holographic image	—
$\hat{Q}$	transformation matrix	—
w, d	width and height of the output	—
P	prediction output	—
T	true output label	—
D	desired intensity	—
H	computed intensity	—
p	amplitude or phase of unit-antenna	—
F	fitness score	—
Y_0	Bessel function	—
Ω	coupling matrix	—
E_{out}	output electric field	—
n_o, n_e	refractive index of ordinary and extraordinary axis	—

Chapter 4

$\vec{x}$	field value	—
$\vec{p}$	parameters	—
f	forward computation	—
g	merit measure	—
λ	adjoint solver	—
F	merit function	—
σ	merit function region	—
p^{ind}	induced polarization	—
α	Claussiu-Mossotti coefficient	—
ϵ_2	permittivity of the inclusion material	—
ϵ_1	permittivity of the original material	—
E^A	adjoint electric field	—
$n(r)$	refractive index distribution	—
g	constant	1.8 mm^{-1}
L	optical path	—
ϕ	phase mask	rad
(x, y, z)	Cartesian coordinate	—
(r, θ, z)	cylindrial coordinate	—
k_0	constant	—
t_0	constant	—
s	constant	—
η	incident angle	50°
$\Delta\rho$	phase delay	rad
d	fabricated depth	nm
λ_i	different wavelength	—
d	length of the design region	—
P	power transmission	—
F	merit function	—