Applications of Metasurfaces in Endoscope and Hyperspectral Imaging

Ningzhi Xie

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Karl F. Böhringer, Chair

Arka Majumdar

Sajjad Moazeni

Program Authorized to Offer Degree:

Electrical and Computer Engineering

University of Washington

Abstract

Applications of Metasurfaces in Endoscope and Hyperspectral Imaging

Ningzhi Xie

Chair of the Supervisory Committee:

Karl F. Böhringer

Electrical and Computer Engineering

Metasurfaces are ultrathin arrays of subwavelength scatterers. They offer versatile control over light within a wavelength-scale thickness. Their ability to condense complex optical functions into compact forms makes them particularly promising for miniaturized imaging systems, where conventional optics face challenges of bulk and limited scalability. This thesis presents several research projects that explore the integration of metasurfaces into endoscopic imaging and hyperspectral imaging, addressing key challenges in device miniaturization, resolution, and spectral functionality.

The first part of the thesis focuses on scanning fiber endoscopes (SFEs), which are

among the most compact scanning-based endoscopes. Two projects demonstrate the replacement of conventional refractive lens assemblies with metasurface-based flat lenses (metalenses). A monochromatic near-infrared metalens was designed, fabricated, and experimentally validated, achieving diffraction-limited performance and significantly reducing optical track length compared to refractive optics. Building on this, a polychromatic metalens was developed to enable tri-color RGB imaging, overcoming the intrinsic dispersion of conventional metalenses and delivering near-diffraction-limited resolution across multiple wavelengths. These studies highlight the potential of metalenses to enable highly compact endoscopic systems with improved imaging performance.

The second part of the thesis extends metasurface functionality beyond focusing to spectral encoding. A metasurface–Fabry–Pérot cavity array was designed as a spatial-to-spectral encoder, enabling the transmission of multi-pixel image information through a single fiber core without scanning. This proof-of-concept demonstrates the feasibility of spectrally encoded, non-scanning endoscopic imaging, offering a pathway to surpass the resolution limits imposed by fiber pixel density in endoscopic imaging.

The final part explores hyperspectral imaging, where metasurfaces are used as spectral code masks for compressive sensing–based reconstruction. A metasurface code mask was optimized to encode full hyperspectral datasets into single-shot grayscale images, which were computationally reconstructed to recover high-resolution hyperspectral information. Experimental demonstrations validate this system as a compact, efficient, and high-speed alternative to conventional hyperspectral imagers.

Together, these works establish metasurfaces as powerful optical platforms for advancing miniaturized endoscopy and hyperspectral imaging, demonstrating the transformative potential of metasurfaces in next-generation biomedical and imaging technologies.

Acknowledgements

I am deeply grateful to all those who have supported me throughout my pursuit of a Ph.D. degree.

First and foremost, I would like to express my sincere appreciation to my advisor, **Prof. Karl Böhringer**. He is a calm and elegant scientist and engineer, and his composure has left a lasting impression on me. Whenever I felt frustrated by setbacks in experiments, he was always there to sit down with me, carefully analyze the problems, and provide valuable insights. His patience and working style have profoundly influenced me, teaching me an essential attitude toward research: when encountering difficulties or unexpected results, remain calm and recognize them as a natural part of scientific discovery.

I am equally indebted to my co-advisor, **Prof. Arka Majumdar**, whose creativity and enthusiasm are boundless. He constantly generates ideas—some brilliantly clever, others boldly unconventional—and all of my research projects have been closely tied to his vision. From him, I have learned that research is ultimately about taking risks and daring to explore the unknown.

Not every minute of conversations with both of my advisors is research related, technical discussions. Prof. Böhringer often began our meetings with casual exchanges, sharing moments of daily life that lightened the mood. Prof. Majumdar, during group meetings, would occasionally share anecdotes from the research community or his own reflections on various topics. These moments offered refreshing breaks from exhaustion and brought renewed energy to my work.

I am also deeply thankful to my colleagues and collaborators. **Johannes Fröch** is a brilliant experimentalist and one of the most generous people in Prof. Majumdar's group. Always ready to help, he could assemble optical setups at least three times faster than the rest of us, and I learned much from him about free-space optical design and alignment. **Zhihao Zhou**, my partner on the computational inverse design of a metalens, impressed me with his strong sense of responsibility. As his first Ph.D. project, we inevitably made some naïve mistakes, and several rounds of experiments did not yield the expected results. Zhihao felt disheartened, but I assured him that this is precisely what research entails. In the end, we finally made it and did a great job.

I also wish to thank **Prof. Eric Seibel**, a pioneer in the field of endoscopy, whose work provided the starting point for many of my projects and whose suggestions were invaluable. **Matthew Carson**, an experienced electrical and mechanical engineer, offered crucial assistance in tuning the fiber scanner for my scanning fiber endoscope work. **Prof. Vishwanath Saragadam** brought his expertise in compressed-sensing computational reconstruction, and our collaboration on the metasurface-enabled hyperspectral imaging project was especially fruitful.

A significant portion of my research was carried out at the **Washington Nanofabrication Facility (WNF)**, where I fabricated all of the metasurfaces used in this thesis. Nanofabrication demands constant alertness—every tiny step matters, and small errors can lead to hours of troubleshooting. The experience was both a rigorous test of my patience and an invaluable training in attention to detail. I am grateful to the dedicated WNF staff who provided technical guidance on specialized tools: **Jean Nielsen** (optical lithography), **Cameron Toskey** (metrology), **Shane Patrick** (electron beam lithography), **Darick Baker and Fred Newman** (deposition), and **Qi Tong**

and Mark Morgan (etching). Their expertise and support were indispensable to my work.

Finally, I want to express my deepest gratitude to my parents. Though they know little about the specifics of my research, they have always been unwavering in their support. Our weekly phone calls, during which they patiently listened to my complaints and frustrations, gave me the courage to persevere. I know you are always there for me so not thing to be afraid of because no setbacks are really big deals.

Contents

List of Figures

List of Tables

List of Abbreviations

SFE	**S**canning **F**iber **E**ndoscope
RGB	**R**ed **G**reen **B**lue
SSE	**S**patial-to-**S**pectral **E**ncoding
FWHM	**F**ull **W**idth at **H**alf **M**aximum
RMS	**R**oot **M**ean **S**quare
RCWA	**R**igorous **C**oupled **W**ave **A**pproximation
NA	**N**umerical **A**pperture
FOM	**F**igure **O**f **M**erit
EBL	**E**lectron **B**eam **L**ithography
DNN	**D**eep **N**eural **N**etwork
SEM	**S**canning **E**lectron **M**icroscope

Chapter 1

Overview

Metasurfaces are ultrathin, engineered two-dimensional arrays of densely packed sub-wavelength optical scatterers, known as meta-atoms. These devices condense complex optical functions into films with just wavelength-scale thickness. One of the major applications is the metalens, a flat optical element capable of replacing conventional refractive lenses, making it an ideal candidate for miniaturized endoscopy. An endoscope is a medical device equipped with a light source and a camera to visualize internal body cavities for diagnostic and therapeutic procedures. As endoscopes continue to shrink in physical dimensions in order to reach more confined spaces like cardiovascular vessels, brain tissue, or spinal regions, conventional optics face limitations in size that metalenses can overcome thanks to their compact form.

Metalenses are inherently wavelength-sensitive: their light manipulation functionalities vary significantly with wavelength, resulting in dispersion, and thus having a narrow operational bandwidth. Additionally, they typically require coherent illumination. Fortunately, endoscopic systems often employ narrow-band laser illumination, which aligns perfectly with these constraints.

Metasurfaces can also produce rich transmission spectra by exploiting resonance modes between neighboring meta-atoms. This pronounced spectral response makes them particularly well-suited for use as spectral filters. Compared with traditional filters based on optical absorption materials or Fabry–Pérot cavities, metasurface-based spectral filters provide distinct advantages, including simplified fabrication and enhanced spectral tunability. These attributes highlight their strong potential for a wide range of spectrum-related applications.

One promising application is spectrally encoded imaging, where spatial intensity information is encoded into the spectrum of light. This approach allows a single fiber core to transmit multiple pixels of image, thereby improving resolution in endoscopic imaging systems that rely on optical fibers for image transfer.

Another application is hyperspectral imaging, which requires capturing the spectrum of every pixel of a scene. In contrast, conventional camera sensor chips are limited to recording grayscale intensity and typically exhibit a broadband response. To achieve hyperspectral functionality, such sensors need to be integrated with spectral filters that selectively transmit light at specific wavelengths.

This thesis demonstrates the adoption of metasurfaces in endoscopes and hyperspectral imagers. Chapter 2 introduces the fundamentals of metasurfaces, covering the basic structure as well as design and simulation theory. Chapter 3 gives an overview of endoscopic imaging. This chapter introduces the basic structures, configuration, and imaging specifications of endoscopes, and discusses the challenges of miniaturization of endoscops. Chapters 4-6 present two projects integrating metalenses into endoscopes, addressing the miniaturization challenges outlined in Chapter 3. Chapter 7 demonstrates spectral encoding of images using metasurface-based spectral filters, which highlights the potential of breaking the resolution limit due to pixel density as discussed in Chapter 3. Chapter 8 demonstrate a project that incoporates metasurface spectral filters in a hyperspectral imager. Chapter 9 outlines the remaining challenges encountered in the research projects presented in this thesis and conceptually introduces ideas of potential approaches to address them.

Chapter 2

Metasurface fundamentals

2.1 Metasurface basic structure

A metasurface consists of a two-dimensional quasi-periodic array of subwavelength scatterers, commonly referred to as meta-atoms, as shown in Figure 2.1a. It can be regarded as the optical analog of a two-dimensional crystalline material, featuring a defined lattice structure. Each lattice site hosts a unit cell, which contains one or multiple meta-atoms. These meta-atoms interact with incident light, enabling modulation of its phase, amplitude, and polarization. Unlike conventional crystalline materials where all unit cells are identical, metasurfaces typically employ spatially varying unit cells. The meta-atoms in these cells maintain a uniform thickness but differ in their lateral dimensions, allowing them to impart position-dependent modulation to the incident light.

2.2 Metasurface functionality, simulation, and design

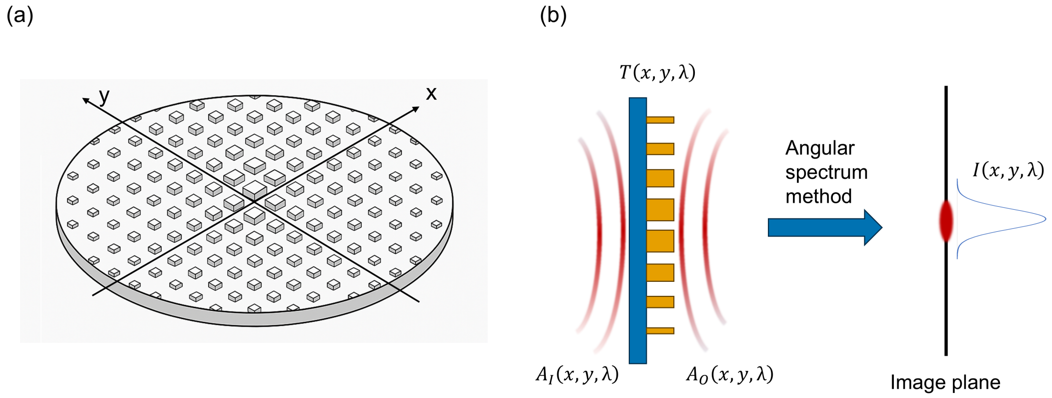

FIGURE 2.1: **(a)** A schematic of a metasurface. **(b)** A schematic of metasurface's functionality and simulation approach.

2.2.1 Metasurface functionality

The metasurface discussed and demonstrated in this thesis is a local metasurface, following the local modulation approximation. This approximation assumes that the modulation of the light field by the metasurface at a specific location (x,y) is solely determined by the unit cell at that location, with negligible influence from neighboring unit cells. The modulation can be polarization-dependent if the meta-atom lacks symmetry along the two orthogonal axes. However, in the research demonstrated in this thesis, we consider only polarization-independent metasurfaces. At the same time, while the modulation may also depend on the angle of incidence, this effect is typically minimal for small angles and is therefore neglected in this thesis. Furthermore, although metasurfaces can be designed to manipulate either reflected or transmitted light (e.g., by incorporating a reflective substrate), this thesis focuses exclusively on transmissive metasurfaces.

As illustrated in Figure 2.1b, to mathematically describe the modulation of the transmitted light, we model the light field as a complex scalar field $A(x, y, \lambda)$ under the scalar diffraction theory. The spatially varying modulation imposed by the metasurface is described by a complex modulation field $T(x, y, \lambda)$, expressed as:

$$T(x,y,\lambda) = t(x,y,\lambda) \cdot e^{i\phi(x,y,\lambda)} \tag{2.1}$$

where t represents the amplitude modulation and ϕ the phase modulation.

The resulting transmitted light field after modulated by the metasurface is given by:

$$A_{\mathrm{O}}(x,y,\lambda) = T(x,y,\lambda)A_{\mathrm{I}}(x,y,\lambda) \tag{2.2}$$

where A_{I} and A_{O} denote the incident and transmitted light field, respectively.

2.2.2 Light propagation simulation

After passing through the metasurface, the modulated light typically propagates through free space or a homogeneous medium before reaching a target plane, where it forms a specific intensity distribution $I(x, y, \lambda)$, as shown in Figure 2.1b. Accurately simulating this light propagation is essential for predicting and designing the optical behavior of the system. An efficient and widely used approach for this purpose is the angular spectrum method. This method decomposes the light field into a superposition of plane waves with different wave vectors, each of which propagates independently according to wave propagation principles.

Mathematically, given a complex light field at $z = z_0$, noted as $A_{z_0}(x, y, \lambda)$, this light field propagates to the plane $z = z_{im}$. On this plane, the light field becomes $A_{z_{im}}(x, y, \lambda)$, resulting in light intensity distribution $I_{im}(x, y, \lambda)$. First, the angular spectrum is obtained by taking the 2D spatial Fourier transform of the light field:

$$U_{z_0}(k_x, k_y, \lambda) = \mathcal{F}\{A_{z_0}(x,y,\lambda)\} \tag{2.3}$$

where (k_x, k_y) are the spatial frequency components, and $U(k_x, k_y)$ represents the amplitude of each plane wave component.

Subsequently, each plane wave component propagates independently and accumulates a phase according to its propagation distance:

$$U_{z_{im}}(k_x, k_y, \lambda) = U_{z_0}(k_x, k_y, \lambda) \cdot e^{ik_z(z_{im}-z_0)} \tag{2.4}$$

where:

$$k_z = \sqrt{k - k_x^2 - k_y^2}\,;\, k = \frac{2\pi}{\lambda} \tag{2.5}$$

The field at the target plane is then calculated by converting the propagated angular spectrum back into the spatial domain:

$$A_{z_{im}}(x, y, \lambda) = \mathcal{F}^{-1}\{U_{z_{im}}(k_x, k_y, \lambda)\} \tag{2.6}$$

The light intensity distribution is calculated from the complex field:

$$I_{im}(x, y, \lambda) = |A_{z_{im}}(x, y, \lambda)|^2 \tag{2.7}$$

2.2.3 Meta-atom response library

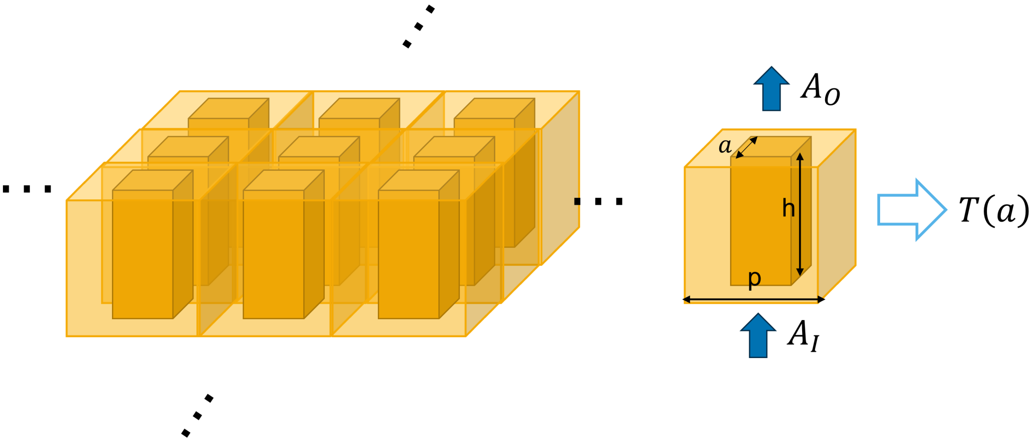

FIGURE 2.2: The setup for obtaining the meta-atom response from RCWA or FDTD simulation.

To design a metasurface with a specific complex modulation field $T(x, y, \lambda)$, we need to determine the complex modulation T as a function of the lateral sizes (termed the geometry parameters) of the meta-atoms. Suppose the meta-atom is defined by a geometry parameter a, then the mapping $a \to T$, denoted as $T(a)$, must be established. Different geometry parameters a lead to different complex modulations T, forming a meta-atom response library. Meta-atoms at each lattice site are selected from this library to realize the desired complex modulation field $T(x, y, \lambda)$, thereby achieving the intended optical functionality.

The meta-atom response library can be constructed through numerical simulations using methods such as rigorous coupled-wave approximation (RCWA) or finite-difference time-domain (FDTD) simulation.

In these simulations, as shown in Figure 2.2, a unit cell is defined with dimensions equal to the metasurface lattice periodicity p. Periodic boundary conditions are applied in the lateral directions, implying that the meta-atom forms part of an infinite periodic array. The simulation setup specifies the refractive indices of the meta-atom and cladding materials, the geometry parameters a, the height of the meta-atom h, and the incident light source—assumed in this thesis to be a normally incident plane wave. The complex fields above and below the meta-atom are monitored, and the complex modulation T (comprising both amplitude and phase components) is extracted by calculating the average complex transmission across the unit cell:

$$T = \frac{\overline{A_{\mathrm{O}}}}{\overline{A_{\mathrm{I}}}} \tag{2.8}$$

where the overline denotes averaging over the entire unit cell area, and A_{I} and A_{O} represent the incident and transmitted complex fields, respectively.

2.2.4 Metasurface design

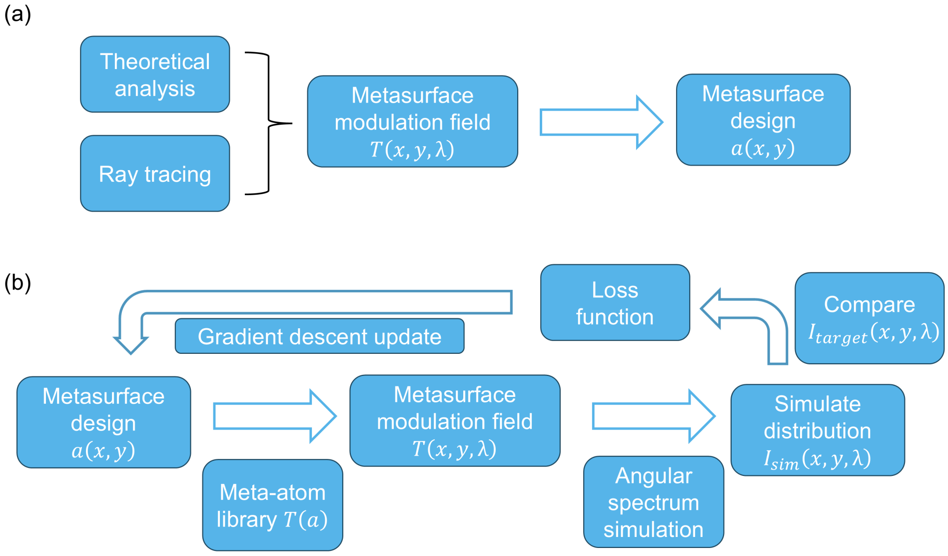

FIGURE 2.3: **(a)** The workflow for designing a metasurface by directly designing the complex modulation function $T(x, y, \lambda)$. **(b)** The workflow for designing a metasurface by iterative optimization.

The functionality of a metasurface can be described as follows: given an input field $A_{\mathrm{I}}(x, y, \lambda)$, the goal is to obtain a target intensity distribution $I_{\mathrm{target}}(x, y, \lambda)$ on the image plane, as illustrated in Figure 2.1b. As shown in Figure 2.3a, to achieve this, the required complex modulation field $T(x, y, \lambda)$ can either be derived from theoretical analysis or obtained through ray tracing simulations. Then, for each spatial location (x, y) and wavelength λ, the appropriate meta-atom that realizes the desired complex modulation T is selected from the meta-atom library.

In some cases, the target modulation field $T(x, y, \lambda)$ cannot be directly determined from theory or ray tracing, or it may not be fully realizable due to physical or fabrication constraints. In such situations, an iterative optimization algorithm can be employed to make the simulated intensity distribution $I_{\mathrm{sim}}(x, y, \lambda)$ match the target distribution $I_{\mathrm{target}}(x, y, \lambda)$ as much as possible. As shown in Figure 2.3b, this process typically involves using the angular spectrum method to simulate the propagation of the modulated output light field and applying a gradient descent algorithm to minimize a loss function that quantifies the difference between $I_{\mathrm{sim}}(x, y, \lambda)$ and $I_{\mathrm{target}}(x, y, \lambda)$.

Chapter 3

Endoscope System Overview

3.1 Introduction to the endoscope

The endoscope is a pivotal medical-imaging instrument that gives clinicians direct visual access to otherwise unreachable hollow organs and body cavities. By delivering illumination and capturing real-time images, it supports both diagnosis and surgery. Endoscopes are generally classified as rigid, flexible, or capsule endoscopes. Flexible endoscopes dominate routine practice because they can navigate tortuous anatomical pathways and readily accommodate working channels for therapeutic tools. Consequently, the discussion in this chapter and the works demonstrated throughout this thesis are mainly based on flexible endoscopic systems.

3.1.1 Structures and configurations

A typical flexible endoscope comprises a miniaturized rigid imaging head at the distal end (inserted into the body) attached to a flexible shaft connecting the distal end and the proximal end (outside the body), as illustrated in Figure 3.1a. The imaging head captures images inside the human body, while the flexible shaft contains a bundle of fibers or fibers plus cables, which delivers the light to the distal end for illumination and relay image data outside of the human body. This flexible shaft allows the instrument to snake through narrow, winding passages with minimal trauma, greatly expanding its clinical utility.

Flexible endoscopes can be broadly categorized into two configurations: non-scanning-based and scanning-based systems. In non-scanning-based endoscopes, as shown in Figure 3.1b, the imaging head functions similarly to a miniaturized camera. The target area is uniformly illuminated by the light delivered to the distal end, and an optical imaging system projects the scene onto an image plane. The resulting image is either transmitted to an external camera via a coherent fiber bundle—a tightly packed array of optical fibers that preserves spatial information—or directly captured by a miniature camera sensor chip placed at the distal tip. In the latter case, the image is converted into electrical signals and transmitted through a cable to external processing equipment.

In contrast, scanning-based endoscopes rely on a different imaging approach. The illumination light is projected by an optical system, typically forming a single spot that moves sequentially to cover the entire scene. As in spot scans, the intensity of the reflected light is recorded as a function of time, enabling the reconstruction of the image pixel by pixel. As demonstrated in Figure 3.1c, scanning can be achieved through several mechanisms: physically moving the light-emitting tip (commonly

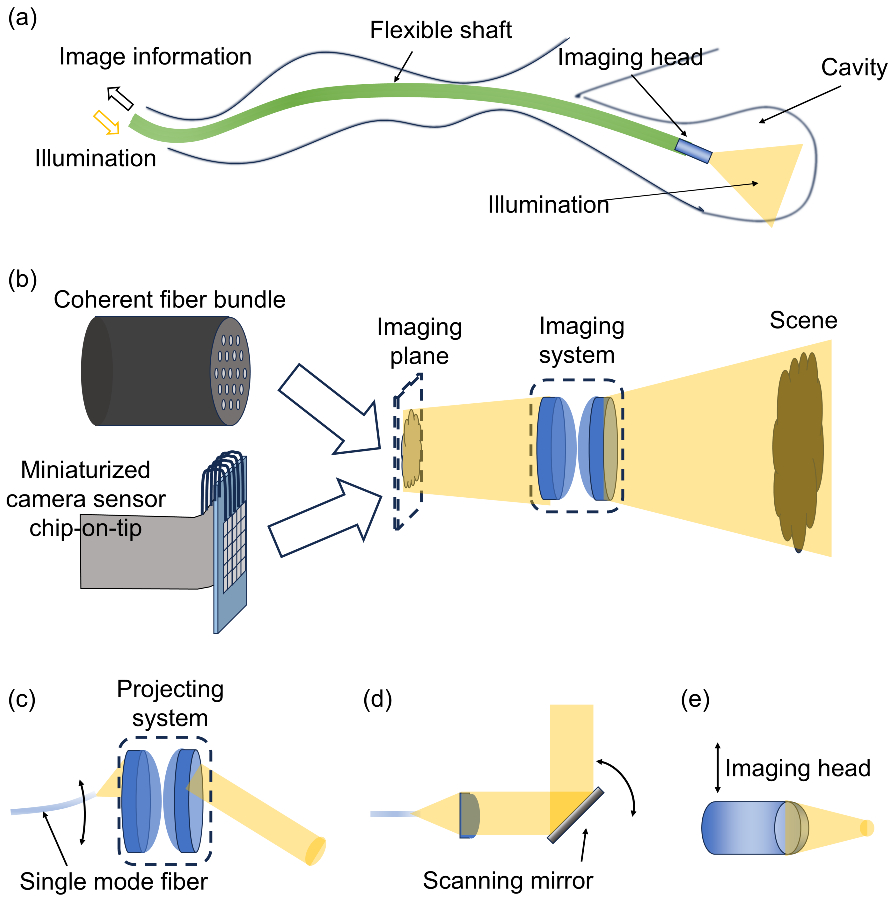

FIGURE 3.1: **(a)** A schematic of a flexible endoscope working inside a body cavity. **(b)** Configurations for non-scanning endoscope. **(c)(d)(e)** Configurations for scanning endoscope, with three major ways for scanning: (c) Moving the light-emitting tip. (d) Scanning mirror. (e) Actuating the entire imaging head.

the tip of an optical fiber) in the imaging head, adjusting the light projecting optical system such as a scanning mirror, or actuating the entire imaging head.

3.1.2 Physical dimensions and imaging specifications

The physical dimensions, the rigid tip length and the outer diameter, are crucial to the application versatility of the endoscope.

As shown in Figure 3.2a, the rigid tip refers to the straight, non-flexible portion at the distal end of an endoscope, typically housing the imaging head and the ending part of the flexible shaft. This length determines how close the tip can approach curved or tortuous anatomy before visualization becomes obstructed. A shorter rigid tip increases maneuverability within narrow ducts and reduces the minimum bend radius required for navigation, which is especially critical in minimally invasive procedures.

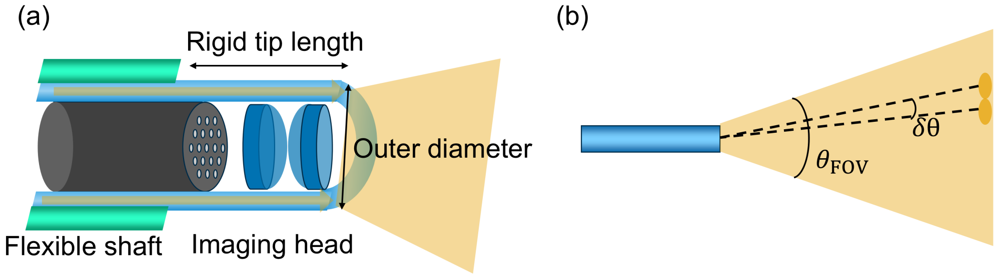

FIGURE 3.2: **(a)** A schematic demonstrating the rigid tip length and outer diameter of an endoscope. **(b)** A schematic demonstrating the field-of-view and angular resolution of an endoscope.

The outer diameter indicates the maximal cross-sectional width of the rigid tip, encompassing the optical system, illumination channels, and any working ports. It directly governs which lumens or anatomical orifices the scope can traverse. Smaller diameters allow access to finer structures and reduce patient trauma, but also constrains the size of internal optics, which affect the imaging specifications.

The key imaging specifications of an endoscope include frame rate, field-of-view (FOV), and spatial resolution, each of which plays a vital role in clinical performance:

Frame rate refers to the number of image frames captured and displayed per second, typically measured in frames per second (fps). A high frame rate enables smooth, real-time video feedback, which is crucial for dynamic procedures and accurate hand–eye coordination during minimally invasive interventions.

As indicated in Figure 3.2b, field-of-view is the angular extent of the observable scene captured by the imaging system, often expressed in degrees. A wide FOV allows clinicians to visualize a larger anatomical area without repositioning the endoscope, enhancing spatial awareness and reducing procedure time. Endoscopes typically have wide field-of-views larger than 70°.

Spatial resolution denotes the minimal lateral separation that the imaging system can resolve. Due to the wide field of view and long imaging distance of the endoscope compared to its diameter, the resolution is usually measured as an angle, as indicated. Higher spatial resolution provides clearer and more detailed visualization of fine anatomical structures, improving diagnostic accuracy and precision during surgical manipulation.

Together, these specifications define the imaging performance of the endoscope and strongly influence its clinical effectiveness.

3.1.3 Applications and needs for miniaturization

Over the past few decades, endoscopes have steadily shrunk in size, opening up new possibilities in exploring tight spaces like the cardiovascular blood vessels [1, 2, 3], the spinal canal, and the deep regions of the brain [4, 5]. Originally, endoscopes were relatively large and rigid, fitting only where direct access was available. But advances in materials, tiny cameras, and light sources have led to ultra-thin, flexible versions that can travel through complex pathways with minimal disturbance. For example, narrow scopes now slip through blood vessels to reveal blockages or guide treatments, while even smaller probes gently navigate around delicate neural tissue deep in the brain or spine. These slimmer instruments are less intrusive, reduce patient discomfort, and allow doctors to examine and sometimes treat areas that were previously unreachable. As miniaturization continues, medical teams can reach more confined spaces, leading to better outcomes in cardiovascular, neurological, and spinal care.

3.2 Challenges and solutions for the miniaturization of the endoscope

There is a clear and accelerating trend toward reducing the physical dimensions of endoscopes, driven by the need to access increasingly confined anatomical regions. However, this miniaturization imposes limitations on imaging specifications and quality. As a result, maintaining high imaging performance while minimizing device size presents a significant technical challenge.

3.2.1 Rigid tip length reduction

In traditional endoscopic imaging heads, the optical imaging or light-projecting system often relies on curved refractive lenses, which contribute significantly to the rigid tip length of the device.

Due to the typically wide field-of-view required in endoscopy, a single spherical lens is insufficient for high quality imaging. At large off-axis angles, such a simple lens would introduce severe optical aberrations including coma, astigmatism, and field curvature that degrade image quality. To mitigate these aberrations, conventional designs use a cascade of multiple curved refractive lenses. These assemblies often involve several lens elements and air gaps, resulting in a longer and more complex optical path that increases the overall length of the endoscopic tip.

To address these limitations, gradient-index (GRIN) lenses have been adopted as a compact alternative to refractive lens assemblies. A single GRIN lens can emulate the function of several discrete lenses by guiding and focusing light through a continuous refractive index gradient within the lens material. Moreover, GRIN lenses feature flat, polished end faces, which simplify alignment and eliminate the need for precise inter-lens spacing. This enables a more compact, rugged, and easily assembled endoscopic tip.

Despite these advantages, GRIN lenses still possess a substantial physical length. In contrast, metalenses offer an ultra-thin solution, with thicknesses on the order of the operating wavelength. These metasurface-based optics can replicate the functionality of complex refractive systems in a single planar element while maintaining low optical aberrations. Their subwavelength-scale phase and amplitude modulation capabilities allow for precise wavefront shaping, making them highly suitable for endoscopic applications where space is constrained.

The integration of metalenses into endoscope designs has already been demonstrated in various configurations, including coherent fiber bundle endoscopes [6, 7], chip-on-tip endoscopes [8], and scanning-based endoscopic systems [9], highlighting their potential to dramatically reduce rigid tip length while preserving or even enhancing imaging performance.

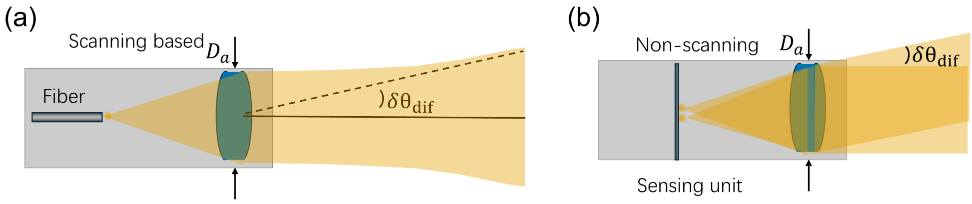

FIGURE 3.3: **(a)** A schematic demonstrating the diffraction limited angular resolution of a scanning based endoscopic system. **(b)** A schematic demonstrating the diffraction limited angular resolution of a non-scanning endoscopic system

3.2.2 Resolution constrained by outer diameter due to light diffraction

As with any conventional optical system, the angular resolution of an endoscope is constrained by the aperture diameter due to the diffraction of light, regardless of whether it is scanning based or non-scanning system. The diffraction-limited angular resolution is given by:

$$\delta\theta_{\text{dif}} = 1.22\frac{\lambda}{D_a} \tag{3.1}$$

where $\delta\theta_{\text{dif}}$ is the diffraction-limited angular resolution, λ is the wavelength of light, and D_a is the aperture diameter of the optical imaging system. In endoscopic systems, D_a is inherently constrained by the outer diameter of the endoscope (D_{OD}), such that $D_a \leq D_{\text{OD}}$.

For scanning-based endoscopes, the illumination beam has a pupil of diameter D_a, determined by either the divergence angle of the light source or the aperture size, whichever is smaller. This finite beam pupil results in a divergence angle that sets the angular resolution, as illustrated in Figure 3.3a.

For non-scanning endoscopes, the incoming light also has a pupil of diameter D_a constrained by the aperture. The collimated beam focused onto the sensing unit produces a diffraction-limited spot size. As shown in Figure 3.3b, two collimated beams with an incident angle difference of $\delta\theta_{\text{dif}}$ form two barely resolvable spots whose separation equals the spot size. This angular difference therefore defines the angular resolution.

In both scanning-based and non-scanning endoscopes, the diffraction-limited angular resolution is governed by the same relation given in Eq. 3.1.

To partially overcome this diffraction limit, post-processing techniques such as image deconvolution can be employed. In particular, the Richardson–Lucy deconvolution algorithm offers a promising approach for enhancing resolution in wide field-of-view endoscopic imaging. Since for such imaging systems, the point spread function (PSF) varies with position across the image field, this algorithm is especially effective because it can accommodate spatially varying PSFs, improving image sharpness beyond the raw optical resolution limit.

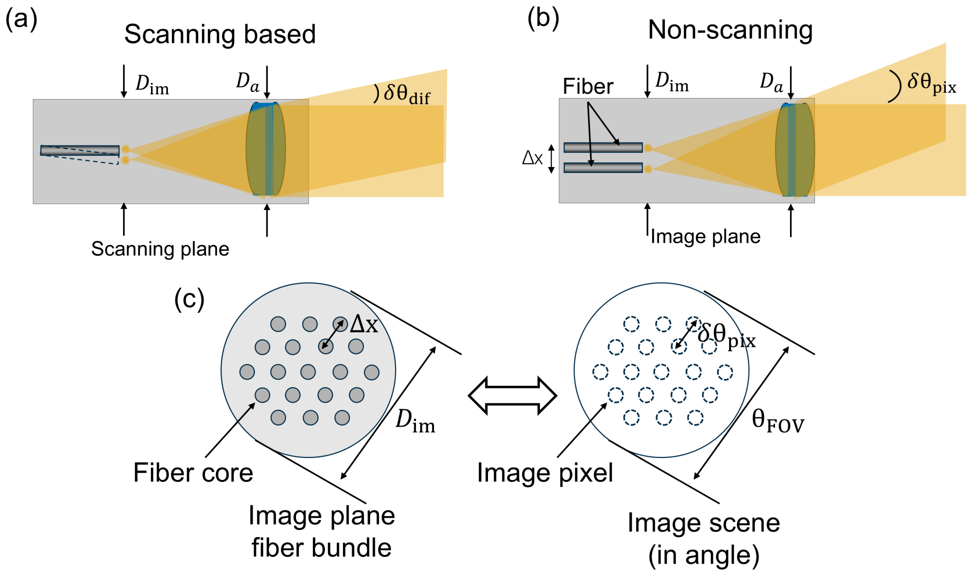

FIGURE 3.4: **(a)** Angular resolution of a scanning based endoscope can reach the diffraction limit. **(b)** A schematic demonstrating the angular resolution of a non-scanning endoscopic system limited by pixel density. **(c)** Finite separation of fiber cores limits the number of pixels of a non-scanning endoscope.

3.2.3 Resolution constrained by outer diameter due to pixel density limitation

For non-scanning endoscopes with a large field-of-view, the diffraction-limited resolution is often unattainable—not due to optical limitations, but due to insufficient pixel density (pixels per unit length) on the image plane. Whether using a coherent fiber bundle or a camera sensor chip-on-tip, the number of resolvable pixels on the image plane is fundamentally constrained by the minimal spacing between adjacent sensing elements (fiber cores or micro sensors), as each sensing element can only capture a single pixel of imaging data. This sets a fundamental ceiling on resolution.

As shown in Figure 3.4b and c, in a coherent fiber bundle, each individual fiber core is separated by cladding and often buffer layers to provide mechanical integrity and to prevent optical crosstalk between neighboring fibers. Similarly, in a CMOS or CCD camera sensor chip, each photosensitive pixel is surrounded by transistors and interconnects necessary for signal readout. These features introduce a minimal center-to-center spacing between pixels, denoted here as Δx, which limits the achievable pixel density on the image plane. Below are the detailed mathematic calculations.

Let the image-capturing area on the image plane have a diameter $D_{\text{im}} \leq D_{\text{OD}}$, where D_{OD} is the outer diameter of the endoscope. The maximum number of resolvable pixels across the diameter is:

$$N_D = \frac{D_{\text{im}}}{\Delta x} \tag{3.2}$$

Given a total field-of-view θ_{FOV}, due to the one-to-one correspondence between the fiber core and the image pixel (see Figure 3.4b), the angular resolution is limited by the pixel density:

$$\delta\theta_{\text{pix}} = \frac{\theta_{\text{FOV}}}{N_D} = \frac{\theta_{\text{FOV}} \cdot \Delta x}{D_{\text{im}}} \tag{3.3}$$

If the spacing between the pixel Δx is significantly larger than the beam spot size on the image plane, the pixel density limited angular resolution $\delta\theta_{\text{pix}}$ will be significantly worse than the diffraction limited angular resolution. In other words, in this case the number of pixels on the image plane is much less than the maximal number of resolvable beam spots. This situation can easily occur in an endoscope, where the diameter of the image-capturing area is small and the angular field-of-view is large.

Assuming the image-capturing area diameter equals the aperture diameter, which is smaller than the outer diameter of the endoscope, i.e., $D_{\text{im}} = D_a < D_{\text{OD}}$, we can compare the pixel density limited resolution $\delta\theta_{\text{pix}}$ to the diffraction-limited resolution $\delta\theta_{\text{dif}}$ from Eq. 3.1. The pixel-limited resolution becomes significantly worse than the diffraction limit when:

$$\theta_{\text{FOV}} \cdot \Delta x \gg 1.22\lambda \tag{3.4}$$

For instance, at a visible wavelength of $\lambda \approx 550\,\text{nm}$, state-of-the-art sensor and fiber technologies achieve a minimum spacing of no better than $\Delta x \approx 3\,\mu\text{m}$. Endoscopic systems typically operate with wide fields of view, often $\theta_{\text{FOV}} \geq 70° \approx 1.22$ radians. In such cases, the inequality in Eq. 3.4 holds true, confirming that pixel density instead of diffraction becomes the dominant limiting factor for angular resolution.

3.2.4 Breaking the pixel density limitation via scanning imaging

For scanning-based endoscopes, each instant of the scan captures a single image pixel, corresponding to a specific position of the light source in the scanning plane. The minimal spacing between light source positions at successive instants defines the total number of pixels of the image and is determined by the signal acquisition rate and the scanning frequency. This spacing can be significantly smaller than the minimal core-to-core spacing in a coherent fiber bundle or the pixel pitch of a camera sensor, enabling sampling at the level required to meet the diffraction-limited angular resolution, as illustrated in Figure 3.4a. Consequently, the pixel count of a scanning-based endoscope, N_D, can greatly exceed that of a coherent fiber bundle or camera sensor, particularly when the image plane diameter D_{im} is small. For example, it has been demonstrated that a scanning fiber endoscope with $D_{\text{im}} = 0.85\,\text{mm}$ can capture 200,000–300,000 total pixels per scene ($\sim N_D^2$), far surpassing the 30,000–60,000 pixels typically achieved by non-scanning endoscopes employing coherent fiber bundles [10].

However, achieving a higher N_D—and thus finer angular resolution—comes at the cost of reduced frame rates and increased system complexity. The need to integrate scanning components adds mechanical and optical intricacies that can significantly increase the rigid tip length, posing additional design and fabrication challenges.

3.2.5 Breaking the pixel density limitation via spectral encoding

A notable property of fiber optics is that light can propagate through the fiber without spectral distortion. This characteristic presents a promising opportunity to overcome the pixel density limitations in endoscopy: by encoding spatial information

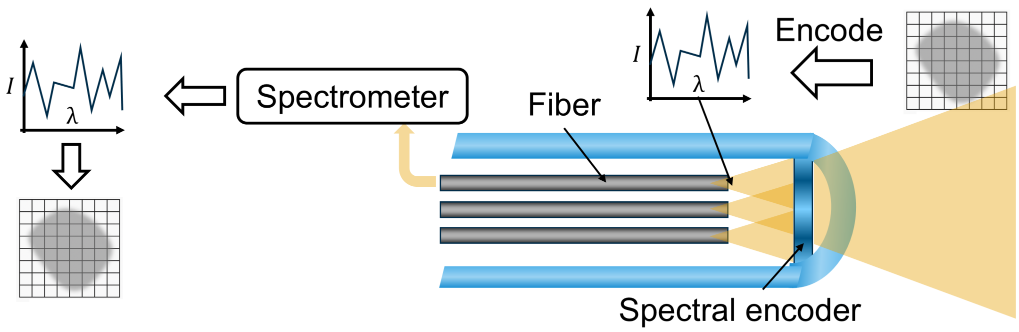

FIGURE 3.5: A schematic demonstrating the spectral encoding approach to break the pixel density limitation.

into the spectrum of light at the distal end using a spectral encoder, and then reconstructing the image at the proximal end, it becomes possible to transmit multiple pixels of information through a single fiber core. Figure 3.5 schematically illustrates this approach.

This approach effectively bypasses the traditional limitations imposed by the minimal core-to-core spacing in coherent fiber bundles. We refer to this concept as Spectrally Encoded Non-Scanning Endoscope (SENSE). The core element of SENSE is the spectral encoder, which has been previously realized using orthogonally tilted Fabry–Pérot cavities [11] or random scattering media [12]. However, these early implementations faced several challenges. First, they relied on inherent randomness in the encoding elements, which limited the ability to optimize the encoding process. Second, they lacked emphasis on microfabrication, resulting in devices that were bulky and unsuitable for integration into compact endoscopic systems.

Metasurfaces, by contrast, offer precise control over the spatial and spectral modulation of light. They are highly designable, allowing for deterministic encoding strategies tailored to specific imaging needs. Moreover, metasurfaces can be fabricated using a single-step lithography process, enabling compact and scalable integration. These advantages make metasurfaces a highly promising solution for implementing SENSE in miniaturized endoscopic platforms.

Chapter 4

Modeling of the scanning fiber endoscope imaging head

4.1 Overview

Scanning fiber endoscopes (SFEs) are a class of scanning-based endoscopes known for their forward-viewing configuration and exceptionally compact size—the smallest among scanning-based designs. In addition to their miniature form factor, SFEs feature a relatively simple optical configuration, making them easier to integrate into medical devices. These advantages make SFEs particularly promising for applications where endoscope miniaturization is critical, such as assisting in clog removal within cardiovascular vessels.

This chapter provides a knowledge base for the following two chapters for two projects related to two research projects in which metasurfaces are employed as flat lenses (metalenses) to replace conventional refractive lens assemblies in SFEs. In the first project, a monochromatic metalens operating at a near-infrared wavelength of 1310 nm was designed, fabricated, and characterized. This wavelength was selected because it lies within a biological "optical window," where bodily fluids such as blood exhibit relatively high transparency. Such reduced optical absorption enhances imaging clarity, making it particularly suitable for applications like angioscopy, where endoscopes are used to provide vision within blood vessels. The design of the metalens follows a direct design approach, in which the metalens is treated as a phase mask with uniform transmission and a spatially varying phase profile. The phase profile is derived from ray-tracing simulations and optimization of the optical system within the endoscope.

The second project focuses on the development of a polychromatic metalens engineered to be dispersionless across multiple discrete wavelengths (441 nm, 532 nm, and 643 nm), enabling tri-color RGB imaging in an SFE. This metalens is designed using an iterative optimization approach: the ideal optical functionality is first determined via ray-tracing simulation, and then the metalens structure is optimized iteratively to closely approximate this ideal performance.

In this chapter, a ray-tracing model was developed to simulate the imaging head of the scanning fiber endoscope, optimize the phase profile for metalens design, and compare the performance of metalens configurations with that of traditional refractive lens assemblies. The simulation results demonstrate that using a metalens can significantly reduce the optical track length. In addition, theoretical calculations of the diffraction-limited resolution were performed. The ray-tracing simulation, along with the resolution analysis, indicates that the metalens configuration can achieve diffraction-limited performance across the entire field-of-view. In contrast, the refractive lens assembly suffers from aberration at large angles, which becomes the dominant factor limiting resolution and causes it to exceed the diffraction limit.

We credit Prof. Eric Seibel and his group members for providing the original ray-tracing model file for refractive lens design of the scanning fiber endoscope.

4.2 Basic structure and working principle of the imaging head

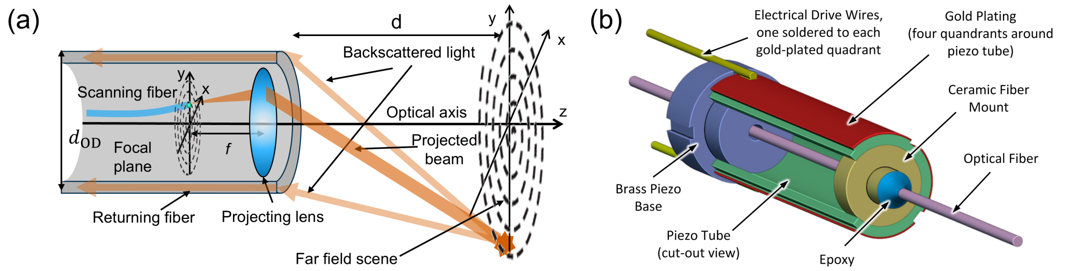

FIGURE 4.1: **(a)** Schematic demonstration the imaging head of a scanning fiber endoscope (SFE). **(b)** Schematic of the structure of the home-made piezo tube actuator for scanning the fiber.

As illustrated in Figure 4.1(a), in the imaging head of an SFE system, a beam is projected to scan the scene at the far field (where the working distance d is much larger than both the diameter of the imaging head D_{OD} and the focal length of the projecting lens f), following a spiral trajectory. As the scanning beam sequentially illuminates different areas of the scene, the backscattered light is collected by a return fiber. The time-varying intensity of the returning light signal is subsequently used to reconstruct the image of the target scene.

The scanning beam is generated by resonantly actuating a single mode fiber. Specifically, the fiber tip follows a spiral trajectory in the focal plane of the metalens, and light emitted from the fiber tip at different lateral positions is projected by different regions of the projecting lens to distinct areas in the far field. This projection can be equivalently viewed as the lens forming an image of the scanning fiber tip at an infinitely distant plane.

The single mode fiber is resonantly actuated using a custom-made piezoelectric tube (Figure 4.1b). The fiber tip position (x, y) follows a spiral trajectory in the focal plane of the metalens, described by:

$$x(t) = A_{\max}(t/T) \cdot \sin \omega t \tag{4.1a}$$
$$y(t) = A_{\max}(t/T) \cdot \cos \omega t \tag{4.1b}$$

Here, T is the period of one complete scan, $t \in [0, T]$, ω is the resonance angular frequency of the scanning fiber system with $\omega T/(2\pi) \gg 1$, and $A_{\max}$ is the maximum lateral displacement of the fiber tip. The beam emitted from (x, y) in the focal plane is projected to a corresponding point (x_I, y_I) in the far field. The time-varying light intensity signal $I(t)$ collected by the return fiber is recorded and synchronized with the scanning trajectory $x_I(t), y_I(t)$ to reconstruct the image signal $I(x_I, y_I)$. The beam scanning trajectory is calibrated prior to imaging.

4.3 Ray-tracing simulation of the beam projection

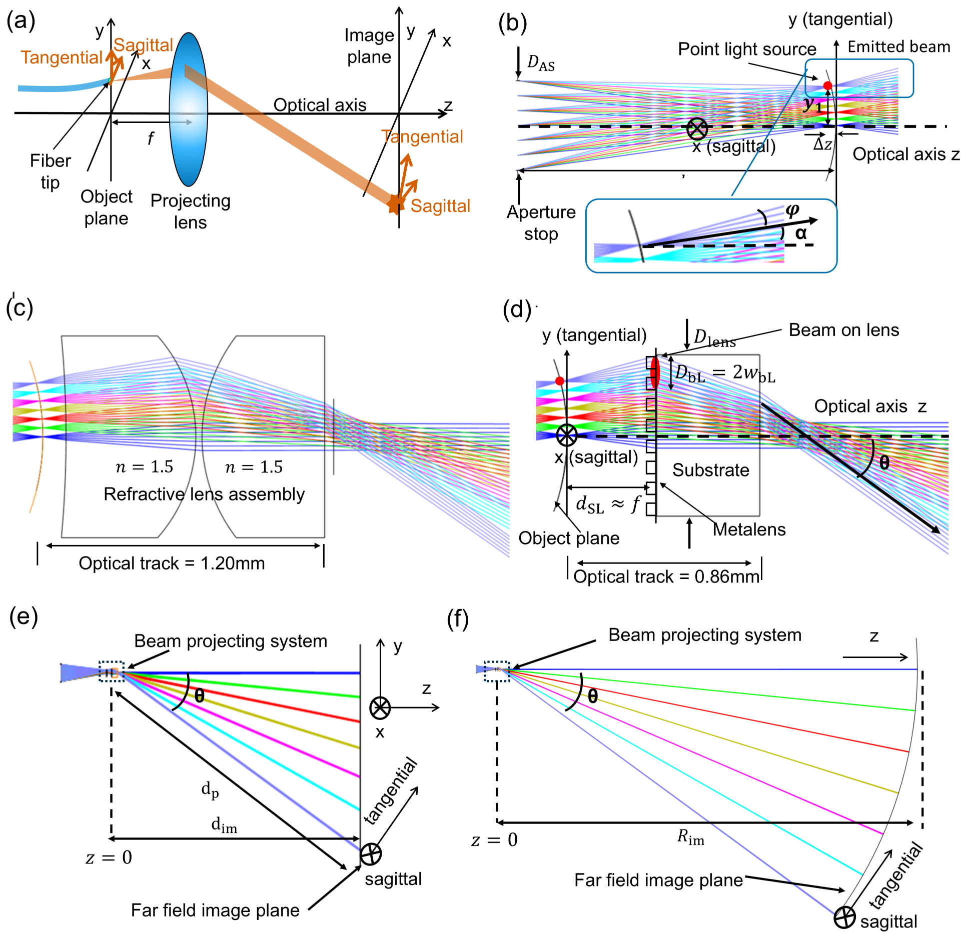

FIGURE 4.2: **(a)** Schematic demonstration of the beam projecting system in SFE. **(b)** Cross-section view of the setup of the light sources imitating a scanning fiber tip in ray-tracing simulation in Zemax. **(c)(d)** Cross-section view of the beam projecting lenses. (c) Traditional refractive lens assembly that consists of two sperical lenses. (d) Metalens acting as a flat phase mask. **(e)(f)** Cross-section demonstration of the beams being projected to different angles at the far field, with image plane imitating the illumination scene. (e) The illumination scene is a plane normal to the optical axis, with a distance d_{im} from the projecting lens. (f) The illumination scene is a spherical surface with its center at the projecting lens position and a radius R_{im}. In this case the projecting beam is studied in angular space.

A ray-tracing simulation of the beam projecting system in the imaging head is constructed in Zemax to guide the design of the metalens used as the projecting lens. As illustrated in Figure 4.2a, the projection system is conceptually divided into three components: the scanning fiber tip (serving as the light emitter), the lens, and the illumination scene. In the simulation, the scanning light emitter is modeled by placing point light sources on the object plane, while the illumination scene is defined as the image plane.

4.3.1 Modeling of the scanning fiber tip

As shown in Figure 4.2b, to emulate the fiber tip that moves along a spiral trajectory, a curved object plane is defined, with multiple light sources placed on this surface.

Since the scanning trajectory—and thus the entire beam projecting system—is rotationally symmetric around the optical axis (z-axis), the light sources positioned along the y-axis at different locations y_1 are sufficient to study the beam projecting system. This arrangement simulates the fiber tip being actuated to varying lateral distances from its resting position on the optical axis. Accordingly, in the object plane, the y-axis is defined as the tangential axis (along the fiber tip lateral displacement direction), while the x-axis becomes the sagittal axis.

As demonstrated in Figure 4.2b, for a fiber tip displaced laterally by y_1, there is a corresponding pointing direction characterized by the beam emitting angle α relative to the optical axis. The fiber tip also exhibits an axial displacement Δz from its resting position. The values of α and Δz as functions of y_1 are determined by the mechanical behavior of the scanning fiber actuated by the piezoelectric tube. These characteristics are modeled using the method described in [13], assuming a 1800 µm free-standing fiber length. The results are summarized in Table 4.1.

The beam emitted from the fiber tip also has a divergence angle φ, which is determined by the numerical aperture (NA) of the fiber:

$$\mathrm{NA}_{\mathrm{fiber}} = \sin \varphi \tag{4.2}$$

In the ray-tracing simulation, we assume $\mathrm{NA}_{\mathrm{fiber}} = 0.18$, which corresponds to $\varphi = 10.4°$.

To approximate the $\Delta z(y_1)$ data in Table 4.1, the object surface is defined as a spherical surface with radius R_{obj}:

$$\Delta z \approx -\frac{1}{2R_{\mathrm{obj}}} \cdot y_1^2 \tag{4.3}$$

From the data in Table 4.1, we obtain $R_{\mathrm{obj}} = 1600\,\mu\mathrm{m}$.

The values of $\alpha(y_1)$ and φ can be modeled by placing an aperture stop to the left of the object plane surface. The parameters are defined by the distance between the aperture stop and the object plane, d_{SS}, and the diameter of the aperture stop, D_{AS}:

$$\alpha \approx \arctan \frac{y_1}{d_{\mathrm{SS}}} \tag{4.4}$$

$$\varphi = \arctan \frac{D_{\mathrm{AS}}}{2d_{\mathrm{SS}}} \tag{4.5}$$

Based on the $\alpha(y_1)$ data in Table 4.1, we determine that $d_{\mathrm{SS}} = 1350\,\mu\mathrm{m}$.

beam	1	2	3	4	5	6	7
fiber tip lateral position y_1 (µm)	0	40	80	120	160	200	240
fiber tip axial displacement Δz (µm)	0	-0.5	-1.5	-4.6	-7.9	-12.4	-17.9
beam emitting angle α (°)	0	1.7	3.4	5.1	6.7	8.4	10.1

TABLE 4.1: The lateral and axial position of the fiber tip and the corresponding beam emitting angles in the ray tracing simulation in Zemax.

4.3.2 Modeling of the projecting lens

As demonstrated in Figure 4.2d, the metasurface serves as a flat beam-projecting lens. As described in Chapter 2.2.1, a metasurface functioning as a metalens can be modeled as a flat phase mask. This phase mask is positioned at an axial distance d_{SL} from the scanning plane of the fiber tip (the object plane), which lies at the focal plane of the lens. The lens forms an image of the fiber tip—treated as a point light source located at the focal plane—onto the far-field scene. To achieve a wide field of view (FOV), the lens must project beams incident near its edge at large off-axis angles. To minimize off-axis aberrations, the lens is initially assigned a primitive phase profile $\phi_0(\rho)$ that approximates a parabolic function:

$$\phi_0(\rho) = -\frac{\pi}{f\lambda}\rho^2 = -\frac{\pi}{d_{\mathrm{SL}}\lambda}\rho^2 \tag{4.6}$$

Here, λ is the wavelength of the light, ρ is the radial distance from the lens center, defined as $\rho(x,y) = \sqrt{x^2+y^2}$, and f is the focal length of the lens.

With this phase profile, the beam projecting angle θ closely follows:

$$\theta = \arcsin\left(y_1/d_{\mathrm{SL}}\right) \tag{4.7}$$

Thus, d_{SL} and the maximum lateral displacement of the fiber tip during scanning, $y_{1\text{-max}}$, determine the system's field of view θ_{FOV} as:

$$\theta_{\mathrm{FOV}} = 2\theta_{\mathrm{max}} = 2\arcsin\left(y_{1\text{-max}}/d_{\mathrm{SL}}\right) \tag{4.8}$$

Given a target $\theta_{\mathrm{FOV}} = 70^\circ$ and $y_{1\text{-max}} = 240\,\mu\mathrm{m}$, as specified in Table 4.1, the required axial distance between the fiber tip and the metalens is $d_{\mathrm{SL}} = 400\,\mu\mathrm{m}$.

The actual phase profile ϕ of the metalens is then slightly optimized to account for the non-flat object plane and the non-zero beam emitting angle α. Due to the rotational symmetry of the projecting lens, the phase profile is expressed as an even-order polynomial:

$$\phi(\rho) = \sum_{k=1}^{n} A_k \rho^{2k} \tag{4.9}$$

Here, k denotes the polynomial order, and A_k are the corresponding coefficients, which are optimized. We set $n = 3$, as the first three terms are sufficient for effective optimization. The initial condition for optimization uses ϕ_0, with $A_1 = -\frac{\pi}{d_{\mathrm{SL}}\lambda}$ and $A_2 = A_3 = 0$.

Figure 4.2c shows a traditional micro-refractive lens assembly with similar functionality, provided here for comparison. Due to fabrication limitations, micro-refractive lenses are typically restricted to spherical surfaces, which do not follow a parabolic phase profile. This results in significant off-axis aberrations, as discussed in Chapter 3.2.1. To mitigate these aberrations, a two-lens spherical assembly is employed. The refractive index of the lens material is approximately 1.5, and the radii and positions of the two lenses are optimized. As shown, compared to a state-of-the-art SFE system using a spherical refractive lens assembly, the metalens configuration reduces the optical track length in the SFE from 1.2 mm to 0.86 mm. This length can be further reduced to approximately 0.4 mm by using a thinner substrate (around 0.1 mm, which is mechanically strong enough to withstand the fabrication process and remain flat in the SFE configuration).

4.3.3 Modeling of the scene being illuminated

The image plane that imitates the scene being illuminated by the beam is either a flat plane normal to the optical axis with work distance $d_{\mathrm{im}} = 15\,\mathrm{mm}$ (Figure 4.2e, for the near-infrared metalens) or a curved spherical surface with large radius R_{im} (Figure 4.2f, for the tri-color RGB metalens). The latter setting is used to study the angular distribution of the projecting beam.

As indicated in Figure 4.2a, e, and f, in the image plane, the saggital direction is still the x-axis and the tangential axis is perpendicular to both the beam propagating direction and the sagittal axis. As will be demonstrated in the following chapters, the intensity profiles of the projecting beams along the sagittal and tangential axes will be different.

4.3.4 Diffraction limited spatial resolution

The spatial resolution of the beam projecting system is determined by both optical aberrations (which is indicated by the spot diagram diameter) and the diffraction-limited diameter of the projected beam. Whichever of the two is significantly larger will dominate and ultimately determine the actual resolution.

The diffraction-limited resolution of this system is defined by the diameter of an ideal collimated Gaussian beam on the image plane. It is not determined by the lens diameter D_{lens}, but instead by an effective aperture, i.e., the beam diameter D_{bL} on the metalens, which is governed by the axial distance d_{SL} and the beam divergence angle φ from the fiber. It can be theoretically calculated as a function of the projecting angle θ and the image plane distance d_{im}.

Assuming the beam emitted from the fiber tip is Gaussian with a divergence angle φ, the beam diameter is twice the Gaussian beam width w_{bL} on the metalens:

$$D_{bL} = 2w_{\mathrm{bL}} = 2d_{\mathrm{SL}} \tan \varphi \tag{4.10}$$

Since the fiber tip is placed at the focal plane of the metalens, the emitted beam becomes collimated upon exiting the lens, with its beam waist located at the metalens. The amplitude profile of this Gaussian beam is assumed to match the incident beam on the lens. Consequently, the projected beam becomes elliptical, with waist widths $w_{0\mathrm{S}}$ and $w_{0\mathrm{T}}$ in the sagittal and tangential directions, respectively:

$$w_{0\mathrm{S}} = w_{\mathrm{bL}} \tag{4.11a}$$

$$w_{0\mathrm{T}} = w_{\mathrm{bL}} \cos \theta \tag{4.11b}$$

As can be seen in Figure 4.2e, the projected beam propagates at an angle θ relative to the optical axis. Therefore, at a working distance z, the actual propagation distance d_{p} is:

$$d_{\mathrm{p}} = z/\cos\theta \tag{4.12}$$

The Gaussian beam width w evolves as a function of the distance d from the waist according to:

$$w(d) = w_0\sqrt{1+(d/z_R)^2} \tag{4.13a}$$

$$z_R = \pi w_0^2/\lambda \tag{4.13b}$$

Substituting $d = d_{\mathrm{p}}$ from Equation 4.8 and using $w_{0\mathrm{S}}$ and $w_{0\mathrm{T}}$ from Equation 4.11, the projected beam widths along the sagittal and tangential directions, $w_{\mathrm{S}}(z,\theta)$ and $w_{\mathrm{T}}(z,\theta)$, are:

$$w_{\mathrm{S}}(z,\theta) = w_{\mathrm{bL}}\sqrt{1+[\lambda z/(\pi w_{\mathrm{bL}}^2\cos\theta)]^2} \tag{4.14a}$$

$$w_{\mathrm{T}}(z,\theta) = w_{\mathrm{bL}}\cos\theta\sqrt{1+[\lambda z\cos\theta/(\pi w_{\mathrm{bL}}^2\cos^3\theta)]^2} \tag{4.14b}$$

This beam width is defined on the plane perpendicular to the propagation direction of the beam. When on the image plane (x-y plane), the beam profile along the y-axis is a projection of the tangential profile. Thus, the beam widths along the x- and y-axes, denoted w_x and w_y, are given by:

$$w_{\mathrm{x}}(z,\theta) = w_{\mathrm{S}}(z,\theta) = w_{\mathrm{bL}}\sqrt{1+[\lambda z/(\pi w_{\mathrm{bL}}^2\cos\theta)]^2} \tag{4.15a}$$

$$w_{\mathrm{y}}(z,\theta) = w_{\mathrm{T}}(z,\theta)/\cos\theta = w_{\mathrm{bL}}\sqrt{1+[\lambda z/(\pi w_{\mathrm{bL}}^2\cos^3\theta)]^2} \tag{4.15b}$$

A Gaussian beam with beam width w has a lateral intensity distribution $I(r)$ described by:

$$I(r) = I_0\exp\left(-2r^2/w^2\right) \tag{4.16}$$

The full width at half maximum (FWHM) of the projected beam on the image plane defines the spatial resolution of the beam projecting system. Accordingly, the resolution $\mathcal{R}$ is:

$$\mathrm{FWHM} = r \text{ where } I(r) = \frac{1}{2}I_0 \tag{4.17a}$$

$$\mathcal{R} = 2\cdot\mathrm{FWHM} = \sqrt{2\ln 2}w \tag{4.17b}$$

Thus:

$$\mathcal{R}_{\mathrm{x}}(z,\theta) = \sqrt{2\ln 2}w_{\mathrm{bL}}\sqrt{1 + [\lambda z/(\pi w_{\mathrm{bL}}^2 \cos\theta)]^2} \tag{4.18a}$$

$$\mathcal{R}_{\mathrm{y}}(z,\theta) = \sqrt{2\ln 2}w_{\mathrm{bL}}\sqrt{1 + [\lambda z/(\pi w_{\mathrm{bL}}^2 \cos^3\theta)]^2} \tag{4.18b}$$

The far-field diffraction-limited resolution can also be defined in angular space, as illustrated in Figure 4.2f. Given the projected beam waist widths $w_{0\mathrm{S}}$ and $w_{0\mathrm{T}}$ in the sagittal and tangential directions, the corresponding angular divergence angles $\delta\theta_{\mathrm{S}}$ and $\delta\theta_{\mathrm{T}}$ are:

$$\delta\theta_{\mathrm{S}} = \lambda/(\pi w_{0\mathrm{S}}) = \frac{\lambda}{\pi w_{\mathrm{bL}}} \tag{4.19a}$$

$$\delta\theta_{\mathrm{T}} = \lambda/(\pi w_{0\mathrm{T}}) = \frac{\lambda}{\pi w_{\mathrm{bL}}} \cdot \frac{1}{\cos\theta} \tag{4.19b}$$

Following the same FWHM-based definition, the angular resolutions in the sagittal and tangential directions, $\mathcal{R}_{\theta\mathrm{S}}$ and $\mathcal{R}_{\theta\mathrm{T}}$, are:

$$\mathcal{R}_{\theta\mathrm{S}} = \sqrt{2\ln 2}\delta\theta\mathrm{S} = \sqrt{2\ln 2}\frac{\lambda}{\pi w_{\mathrm{bL}}} \tag{4.20a}$$

$$\mathcal{R}_{\theta\mathrm{T}} = \sqrt{2\ln 2}\delta\theta\mathrm{T} = \frac{\sqrt{2\ln 2}}{\cos\theta}\frac{\lambda}{\pi w_{\mathrm{bL}}} \tag{4.20b}$$

4.3.5 Ray-tracing results

As shown in Figure 4.2c and d, seven point sources are evenly distributed along the y-axis from $y = 0\,\mu\mathrm{m}$ to $y = 240\,\mu\mathrm{m}$ on the object plane, which is located $400\,\mu\mathrm{m}$ away from the metalens. The phase profile of the metalens is optimized to minimize the root-mean-square (rms) radius of the spot diagrams corresponding to the images of these seven point sources. The spot diagrams of the projected beams for the seven positions are presented in Figure 4.3a–c.

As shown in Figure 4.3d, the beam projecting angle θ as a function of the fiber tip's lateral position y_1 closely follows the theoretical relationship $\theta = \arcsin y_1/d_{\mathrm{SL}}$ (Equation 4.8), for both the refractive lens assembly and the metalens configurations.

The diffraction-limited resolutions (FWHM beam diameters) are calculated using Equations 4.18 and 4.20, assuming $w_{\mathrm{bL}} = 73\,\mu\mathrm{m}$ ($\mathrm{NA}_{\mathrm{fiber}} = 0.18$, $d_{\mathrm{SL}} = 400\,\mu\mathrm{m}$), and $\lambda = 1.31\,\mu\mathrm{m}$ for the near-infrared SFE (Figure 4.3a and b), and $\lambda = 0.441\,\mu\mathrm{m}$ for the tri-color RGB SFE (using the shortest working wavelength, Figure 4.3c). The calculated FWHM beam edges are plotted as orange dashed lines in the spot diagrams.

The diffraction-limited beam diameters are also compared with the diameters of the ray-tracing spot diagrams (defined as two times the rms radius) in Figures 4.3e and f. As observed, the ray-tracing spot diameters for the metalens configurations remain well below the corresponding diffraction-limited diameters, indicating that diffraction dominates and thus defines the system resolution. In contrast, for the refractive lens assembly, the spot diameter exceeds the diffraction limit at large beam projecting angles ($\theta \geq 30°$). This demonstrates that the refractive lens assembly—despite

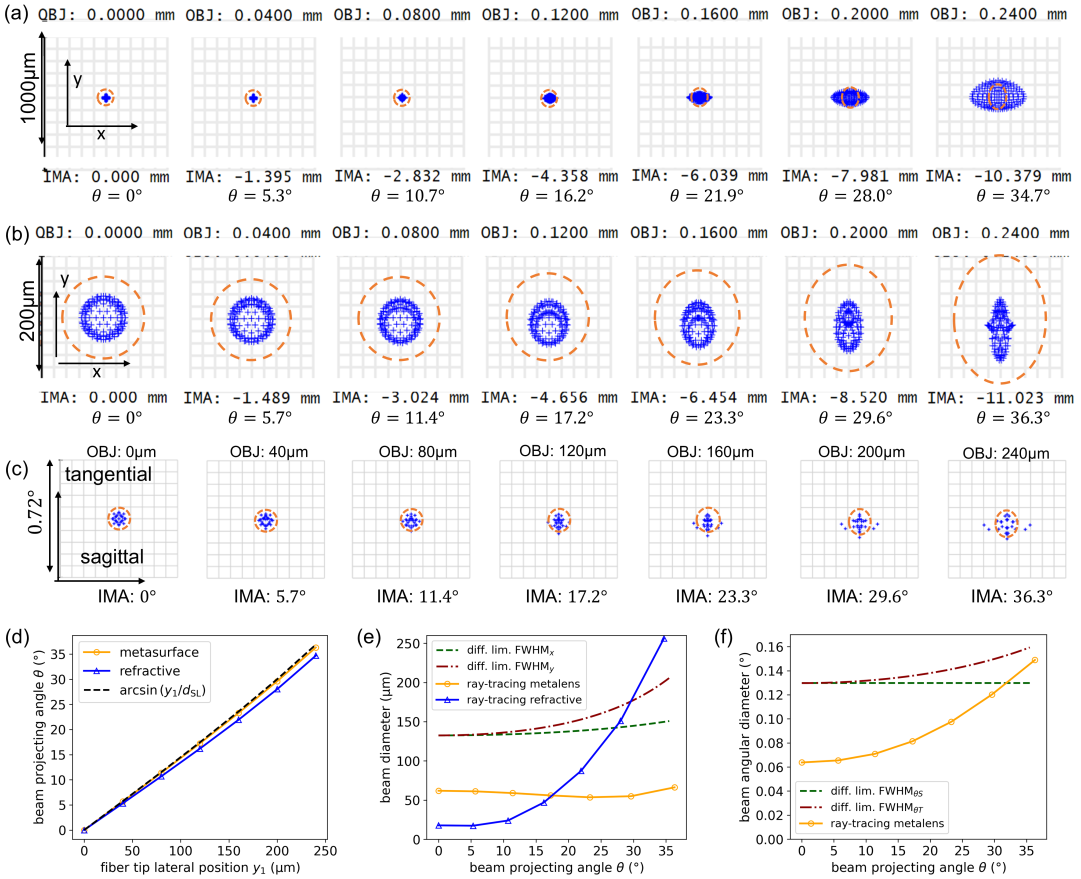

FIGURE 4.3: **(a)(b)(c)** Ray-tracing spot diagram of the SFE beam projecting system after ray-tracing optimization in Zemax. The orange dash-line indicates the diffraction limited FWHM beam edge. **(a)** Refractive lens assembly setup in Figure 4.2c, working wavelength $\lambda = 1.31\,\mu\text{m}$. The image plane is indicated in Figure 4.2e with $d_{\text{im}} = 15\,\text{mm}$. **(b)** Metalens setup in Figure 4.2d, $\lambda = 1.31\,\mu\text{m}$. The image plane is indicated in Figure 4.2e with $d_{\text{im}} = 15\,\text{mm}$. **(c)** Metalens setup in Figure 4.2c, $\lambda = 0.441\,\mu\text{m}$. The image plane is indicated in Figure 4.2f with $R_{\text{im}} = 80\,\text{mm}$. **(d)** Beam projecting angle θ as a function of the lateral displacement of the fiber tip during scanning, y_1. **(e)(f)** Diffraction limited FWHM beam diameter and ray-tracing spot diagram root-mean-square diameter. (e) The cases in (a) and (b). **(f)** The case in (c).

its longer optical track length and more complex configuration—fails to achieve diffraction-limited performance at large θ due to optical aberrations.

These results suggest that replacing the refractive lens assembly in SFE systems with a metalens not only reduces the length of the imaging head, but also improves the overall optical performance.

Chapter 5

Monochromatic metalens for near-infrared scanning fiber endoscope

5.1 Overview

A monochromatic metalens was designed to replace the refractive lens assembly in a near-infrared scanning fiber endoscope. The metalens is specified to have a field-of-view of $\theta_{\mathrm{FOV}} = 70^\circ$, a diameter of $D_{\mathrm{lens}} = 0.68\,\mathrm{mm}$, a working distance of $d_{\mathrm{im}} = 15\,\mathrm{mm}$, and an operating wavelength of $\lambda = 1.31\,\mu\mathrm{m}$. Following fabrication, the metalens was experimentally characterized by measuring the longitudinal and transverse intensity distributions of the projected beam. The results confirm that the metalens achieves diffraction-limited resolution.

We credit Matthew D. Carson for building the fiber actuator that performs fiber scanning.

This chapter is adapted from a published paper [14].

5.2 Metalens design

The ray-tracing optimization is performed as described in Chapter 4.2. Subsequently, the metalens is directly designed to implement the phase profile $\phi(\rho)$ (shown in Figure 5.1a) obtained from ray-tracing optimization. Since the meta-atoms can only realize phase delays on the order of several multiples of 2π, the continuous phase profile is wrapped modulo 2π and then discretized into 12 uniform phase levels, as illustrated in Figure 5.1b. For each lattice site (x, y) on the metalens, the required phase is calculated as $\phi(\rho = \sqrt{x^2 + y^2})$ and then rounded to the nearest discretized level. A corresponding meta-atom is selected from the meta-atom library to realize each phase level.

The meta-atom library is based on crystalline silicon (cSi) square posts on a sapphire (Al_2O_3) substrate, arranged on a square lattice with a periodicity of $p = 600\,\mathrm{nm}$ and a post height of $h = 1000\,\mathrm{nm}$. The lateral size a of the meta-atom is the tunable parameter, as shown in Figure 5.2a. Rigorous coupled wave analysis (RCWA) [15] is used to numerically calculate the phase and transmission responses of the meta-atoms as a function of a. These simulation results guide the selection of meta-atom sizes that achieve high transmission while covering the 12 discretized phase levels required to construct the metalens. The simulated phase and transmission responses are plotted in Figure 5.2b.

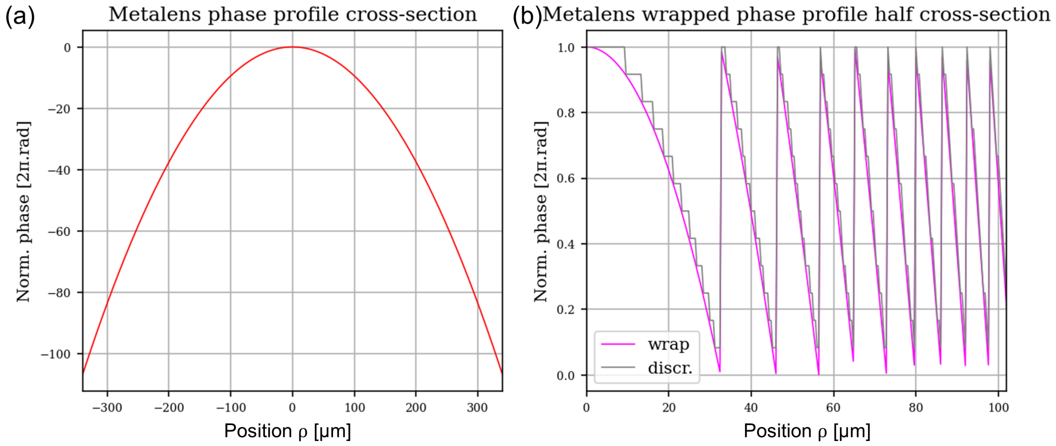

FIGURE 5.1: **(a)** The unwrapped phase profile of the metalens. **(b)** The wrapped and discretized (12 equally spaced levels) phase profile ($0 < \rho < 60\,\mu\text{m}$) of the metalens.

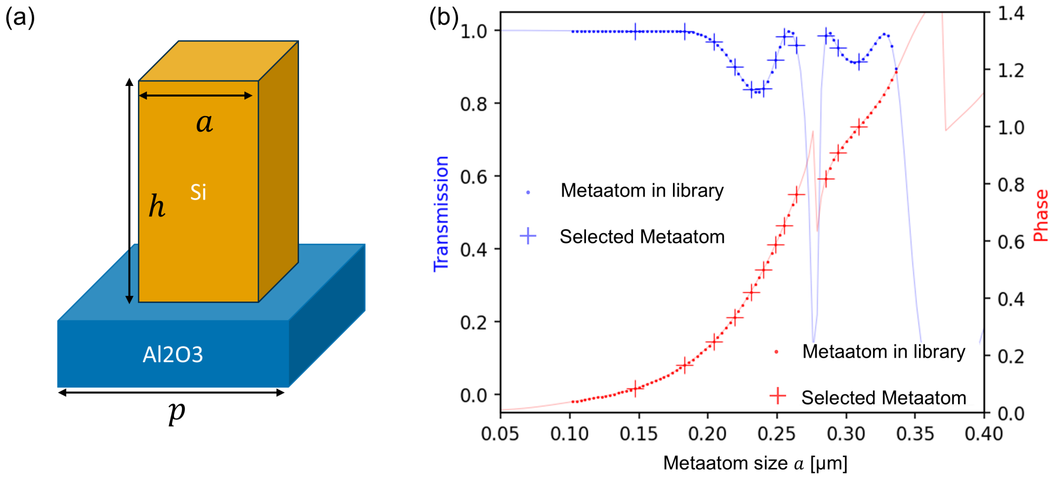

FIGURE 5.2: **(a)** The unit cell of the meta-atom on Al_2O_3 substrate. The height of the meta-atom is $h = 1000\,\text{nm}$ and the periodicity of the meta-atom is $p = 600\,\text{nm}$. **(b)** The wrapped and discretized (12 equally spaced levels) phase profile ($0 < \rho < 100\,\mu\text{m}$) of the metalens.

5.3 Metalens fabrication

The designed metalens was fabricated using electron beam lithography (EBL). The fabrication process flow is illustrated in Figure 5.3a. A layer of electron-beam resist (ZEP-520A) was spin-coated onto a crystalline silicon-on-sapphire substrate with a silicon thickness of 1000 nm and patterned using EBL. A $\sim 50\,\text{nm}$-thick Al_2O_3 hard mask was subsequently deposited by electron-beam-assisted evaporation, followed by lift-off of the resist layer (approximately 200 nm thick) in N-methyl-2-pyrrolidone (NMP) at 90°C overnight. The exposed crystalline silicon layer was then etched using a fluorine-based reactive ion etching process.

Optical microscope images of the fabricated metalens are shown in Figure 5.3b, while scanning electron microscope (SEM) images are presented in Figure 5.3c and d.

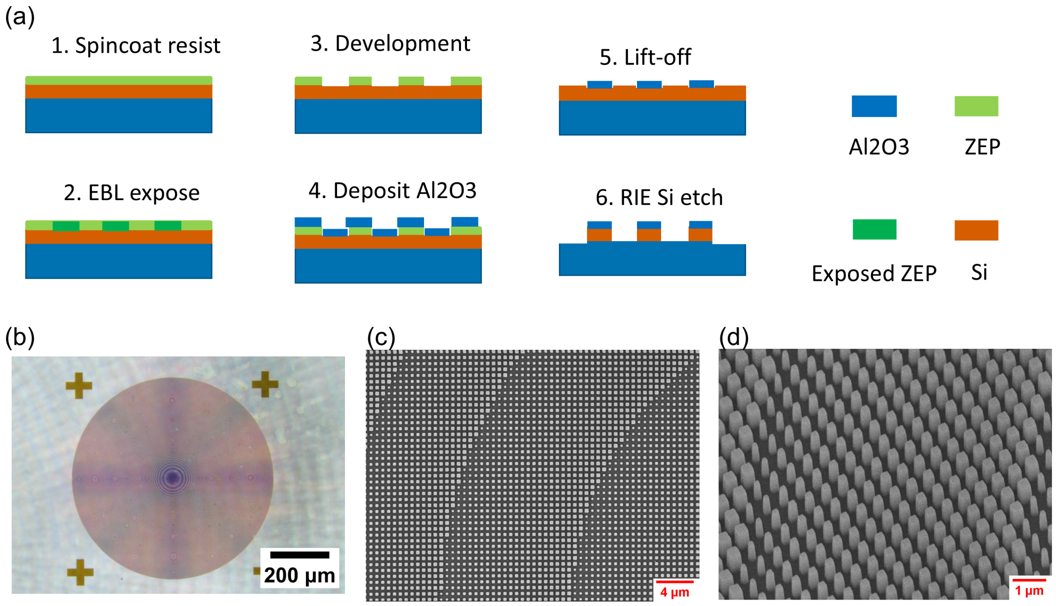

FIGURE 5.3: **(a)** The fabrication process flow of the metalens. **(b)** Optical microscope image of the fabricated metalens. **(c)** Scanning electron microscope (SEM) image of the fabricated meta-lens from the top. **(d)** SEM image at the oblique angle of 40°.

5.4 Metalens characterization

5.4.1 Longitudinal beam intensity distribution

We first characterize the longitudinal intensity distribution of the beam projected by the metalens. Figures 5.4a and b illustrate the schematic of the experimental setup used for this measurement. A movable infrared (IR) camera, consisting of an objective lens (Mitutoyo Plan Apo NIR 10×, $f = 20\,\mathrm{mm}$, $NA = 0.26$), a tube lens (Thorlabs AC254-075-B-ML, $f = 75\,\mathrm{mm}$), and an IR sensor (WiDy SenS S320 V-ST), is positioned along the optical axis in front of the metalens. This configuration enables the characterization of the beam profile on the imaging plane of the camera, indicated by the vertical dashed line in Figures 5.4a and b.

A single-mode fiber mounted on a piezo tube is placed behind the metalens. The piezo tube is affixed to a 3D translation stage, allowing movement along both the optical axis (z-direction) and the transverse (x–y) plane. The fiber is first aligned to the correct lateral position relative to the metalens. By imaging the metalens with the camera, as shown in Figure 5.4a, the fiber can be adjusted in the x–y plane and centered on the metalens. The camera is then positioned at $z = 15\,\mathrm{mm}$ from the metalens, and the fiber is translated along the z-axis to identify the position where the captured beam reaches its minimum size. This corresponds to the optimal source-to-lens distance d_{SL} as designed.

With the fiber fixed at the neutral position ($\theta = 0$), the beam's transverse intensity profile is measured at various axial distances z by moving the camera along the

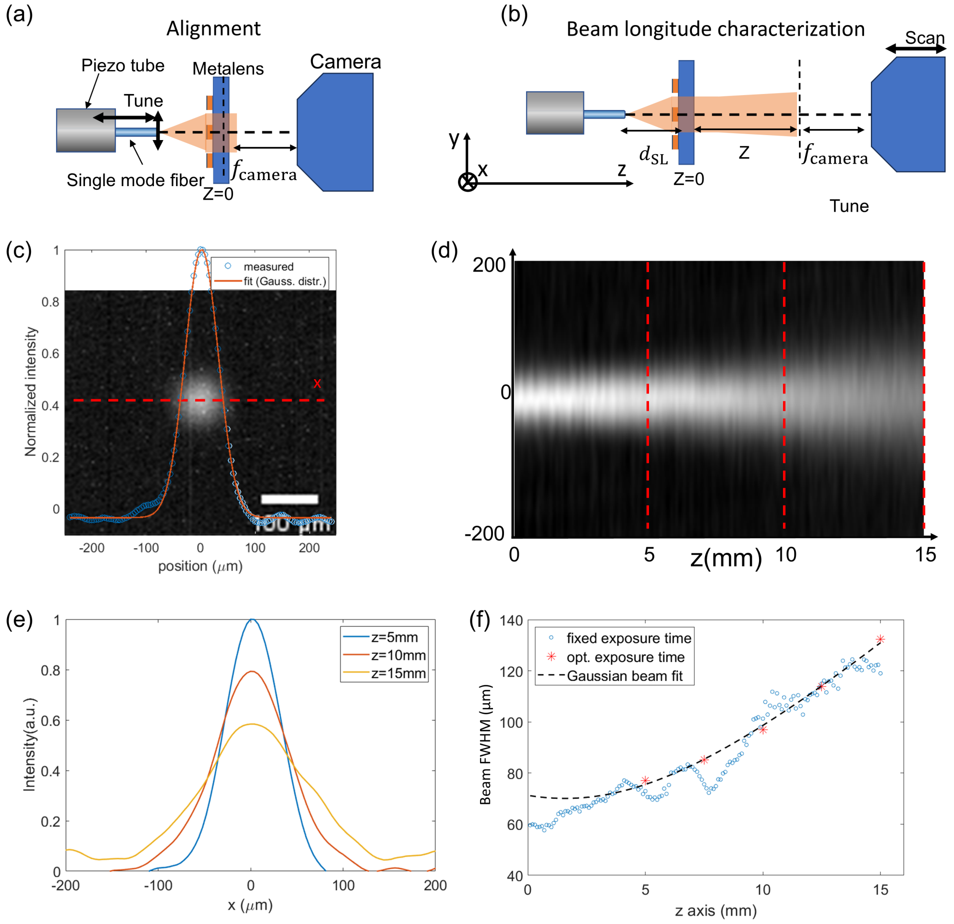

FIGURE 5.4: **(a)** Schematic of the alignment of fiber to the metalens. **(b)** Schematic of the measurement setup for characterizing the longitudinal (y-z plane) beam intensity distribution. **(c)** Direct image of the beam without fiber scanning at the working distance $z = 5$ mm. A Gaussian blur (radius = 12 µm) is applied to the image to reduce the noise. After that, the intensity distribution along the x-axis (the red dashed line) is subtracted and fitted with a Gaussian distribution function. The measured and fitted intensity distribution is plotted on the image. **(d)** Longitudinal beam intensity distribution, constructed from a series of images similar to (c) taken at z ranging from 0 to 15 mm with an increment of 0.1 mm. The beam power from the fiber and the exposure time are fixed. **(e)** Beam intensity distribution along x-axis for various working distances z, corresponding to the three red dashed lines in (d). **(f)** FWHM of the beam intensity along x-axis vs. z at the center of the field. Blue circles correspond to a series of images with a fixed exposure condition. Red stars correspond to images with dynamically optimized exposure conditions to maximize the signal-to-noise ratio. The dashed line is a fit to the red stars, assuming the beam is a perfect Gaussian beam.

optical axis, as indicated in Figure 5.4. These measurements are used to reconstruct the longitudinal beam intensity distribution.

Figure 5.4c shows a representative beam image captured at $z = 5$ mm. The one-dimensional intensity distribution along the x-axis (marked by the red dashed line) is extracted and fitted with a Gaussian profile. Repeating this procedure at multiple z-positions from $z = 0$ to 15 mm allows construction of the longitudinal beam profile in the x–z plane, as shown in Figure 5.4d. These results indicate that the beam diverges only slightly over a propagation distance of 15 mm. Figure 5.4e plots the 1D intensity distributions at different z positions, demonstrating that the beam shape is well preserved with only slight broadening between $z = 5$ mm and $z = 15$ mm.

The full width at half maximum (FWHM) diameter of the beam at $\theta = 0$ is plotted as a function of z in Figure 5.4f. This data is fitted using a modified version of Equation 4.18a, assuming $\theta = 0$ and allowing the beam waist to be located at a position $z = z_0$ (as a fitting parameter) rather than at $z = 0$. The fitting equation is given by:

$$\mathrm{FWHM_x} = \sqrt{2\ln 2} \cdot w_0 \sqrt{1 + \left[\frac{\lambda(z - z_0)}{\pi w_0^2}\right]^2} \tag{5.1}$$

Here, w_0 is the beam waist radius at $z = z_0$, and both w_0 and z_0 are fitting parameters. From the fit, we estimate the beam waist to be located at $z_0 = 1.58\,\mathrm{mm}$, with a waist radius of $w_0 = 59.5\,\mu\mathrm{m}$, corresponding to a FWHM of 70.1 µm. Further along the optical axis, at $z = 15\,\mathrm{mm}$, the beam expands to a FWHM of 135 µm. These results demonstrate that the beam projected by the metalens exhibits a longitudinal intensity distribution closely resembling that of a Gaussian beam, with the beam waist located near the metalens, confirming the theoretical calculation of the diffraction limited resolution of the SFE imaging head in Chapter 4.3.4.

5.4.2 Lateral scanning trajectory

To further characterize the beam projected by the metalens at nonzero projecting angles ($\theta > 0$), we examine the scanning trajectory of the beam as the fiber is actuated. The fiber scanner is driven at its resonance frequency of 2187 Hz using a stable sinusoidal input, producing an elliptical scanning trajectory. The resulting time-averaged beam trajectory imaged on an IR fluorescence card is shown in Figure 5.5a. We take images of this trajectory with a camera exposure time exceeding 10 times the scanning period.

By moving the camera in the x–y plane, we measure the intensity distributions of the scanning beam at different segments of the trajectory, corresponding to varying projection angles, as shown in Figure 5.5b. For large projection angles ($\theta > 14^\circ$), the numerical aperture (NA) of the camera objective is insufficient to directly capture the beam. To address this, a piece of frosted paper is placed at the imaging plane to scatter the beam, allowing the scattered light to be recorded by the camera.

Figure 5.5c presents a close-up of the beam trajectory near the center of the ellipse (corresponding to $\theta = 6.8^\circ$) captured at $z = 15\,\mathrm{mm}$ with an exposure time of 8 ms. Comparison with images captured without the frosted paper reveals that the scattering introduces slight broadening in the transverse profile of the beam trajectory, increasing the full width at half maximum (FWHM) by approximately 15 µm. This effect is corrected in post-processing to recover the intrinsic beam trajectory width.

As illustrated in Figure 5.5d, different points along the elliptical scanning trajectory correspond to different projection angles θ. To quantify how the projected beam diameter varies with θ, we analyze the FWHM of the scanning trajectory at various points along the ellipse. The transverse width of the beam trajectory reflects the diameter of the projected beam, which is directionally dependent.

Notably, for $\theta > 0^\circ$, the intensity profile of the projected beam becomes elliptical. The intensity distributions along the x and y axes differ, as discussed in Chapter 4.3.4. The y-axis is defined as the direction connecting the beam spot to the center point (i.e., the intersection of the optical axis with the x–y plane). The width of the

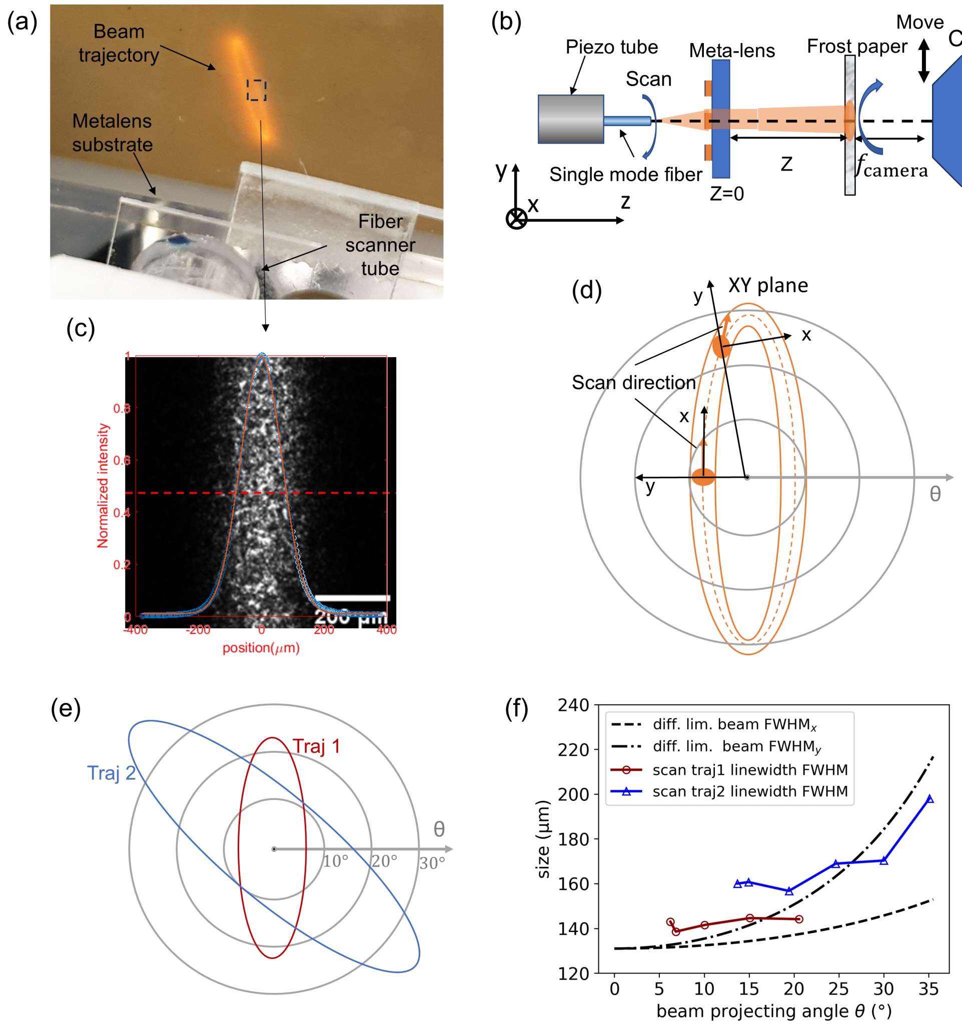

FIGURE 5.5: **(a)** Image of the elliptical trajectory of the beam projected by the meta-lens onto an IR card. The fiber is driven by a piezo tube at the fiber's mechanical resonance frequency. We credit Matthew D. Carson for building the fiber actuator. **(b)** Schematic of the measurement setup for characterizing the beam scanning trajectory on the imaging plane. **(c)** The center part (6.8° beam projection angle, indicated by the dash-line square in (a)) of the beam trajectory with fiber actuation, captured by the IR camera at the distance $z = 15\,\text{mm}$. The left is a direct image, the right is taken with a thin frost paper at the imaging plane of the camera. The intensity distribution across the beam trajectory and averaging along the trajectory is measured and plotted. Gaussian blur (radius = 24 μm for image with the paper) is applied to reduce the noise. **(d)** A schematic showing the beam profiles at different projecting angles θ, and their relationships to the trajectory linewidth. **(e)** Demonstration of the beam scanning trajectories on the imaging plane. **(f)** The experimentally measured and diffraction limited FWHM of the projected beam v.s. the projection angle at $z = 15\,\text{mm}$.

trajectory is influenced by the beam diameter along x and y directions, depending on the segment of the ellipse. As shown in Figure 5.5d, the intensity profile of the trajectory at the co-vertex (minimum θ) is determined by the beam's y-direction profile, while near the vertex (maximum θ), it primarily reflects the x-direction profile.

Two scans were performed using elliptical trajectories with different angular spans. The first trajectory spans $\theta = 6.2^\circ$ to 20.6° (red in Figure 5.5e), and the second spans $\theta = 13.7^\circ$ to 35.0° (blue). The FWHMs of the intensity across the beam trajectory are plotted as a function of projection angle in Figure 5.5f. These measurements are compared against theoretical diffraction-limited FWHMs of the beam derived from modified versions of Equation 4.18:

$$\mathrm{FWHM_x}(\theta) = \sqrt{2\ln 2} \cdot w_0 \sqrt{1 + \left[\frac{\lambda(z - z_0)}{\pi w_0^2 \cos\theta}\right]^2} \tag{5.2a}$$

$$\mathrm{FWHM_y}(\theta) = \sqrt{2\ln 2} \cdot w_0 \sqrt{1 + \left[\frac{\lambda(z - z_0)}{\pi w_0^2 \cos^3\theta}\right]^2} \tag{5.2b}$$

Here, $z_0 = 1.58\,\mathrm{mm}$ and $w_0 = 59.5\,\mu\mathrm{m}$, as determined from the fitting in Figure 5.4f.

As seen in Figure 5.5f, for both small and large field-of-view scans, the measured FWHMs of the trajectory at the co-vertex (minimum θ) agree closely with the diffraction-limited beam FWHM along the y-direction. Likewise, at the vertex (maximum θ), the measured widths match the diffraction-limited beam FWHM along the x-direction. This trend is consistent with theoretical predictions.

In summary, the measured FWHMs of the beam trajectories within the projecting angle range $\theta \leq 35°$ agree well with the theoretical diffraction limits. This confirms that the imaging head incorporating the metalens maintains diffraction-limited resolution across the full design field-of-view, $\theta_{\mathrm{FOV}} = 2\theta_{\max} = 70°$.

Chapter 6

Polychromatic metalens for tri-color RGB scanning fiber endoscope

6.1 Overview

A polychromatic metalens, designed to be dispersionless at three distinct wavelengths, was designed for a tri-color RGB scanning fiber endoscope (SFE) to address the intrinsic dispersion challenge commonly associated with metalenses. Unlike the monochromatic near-infrared metalens, this polychromatic metalens is designed using an inverse-design approach, in which the target optical functionality is first specified from the ray-tracing simulation, and the metalens structure is then iteratively optimized to approximate the desired performance. The metalens shares the same field-of-view of $\theta_{\text{FOV}} = 70^\circ$ and diameter of $D_{\text{lens}} = 0.68\,\text{mm}$ as the monochromatic near-infrared metalens. It simultaneously operates at three wavelengths: $\lambda_{\text{R}} = 643\,\text{nm}$, $\lambda_{\text{G}} = 532\,\text{nm}$, and $\lambda_{\text{B}} = 441\,\text{nm}$. The spatial resolution of the RGB SFE imaging head is characterized in angular space and achieves near-diffraction-limited angular resolution. The ability to maintain achromatic performance across multiple wavelengths comes at the cost of reduced optical efficiency. The average relative efficiency across the three design wavelengths is approximately 32% on-axis, and 13% when averaged over the entire field-of-view.

We credit Zhihao Zhou for performing the inverse design of the polychromatic metalens, and VerAvanti Inc. for providing the SFE test platform.

This chapter is adapted from a published paper [16].

6.2 Addressing the dispersion of metalens

Metalenses, like other diffractive optical elements, exhibit strong wavelength-dependent dispersion, with their focal length inversely proportional to the wavelength [17]. This intrinsic chromatic dispersion introduces significant aberrations, posing a major limitation for multi-color imaging applications. To mitigate chromatic aberrations, meta-atoms can be engineered such that the phase modulation they impart scales linearly with wavelength [18, 19, 20, 21, 22]. This enables the design of phase profiles that focus multiple wavelengths onto a common focal plane. However, this strategy is feasible primarily for metalenses with unwrapped phase profiles (having only one Fresnel zone), which is typically limited to low numerical aperture (NA)

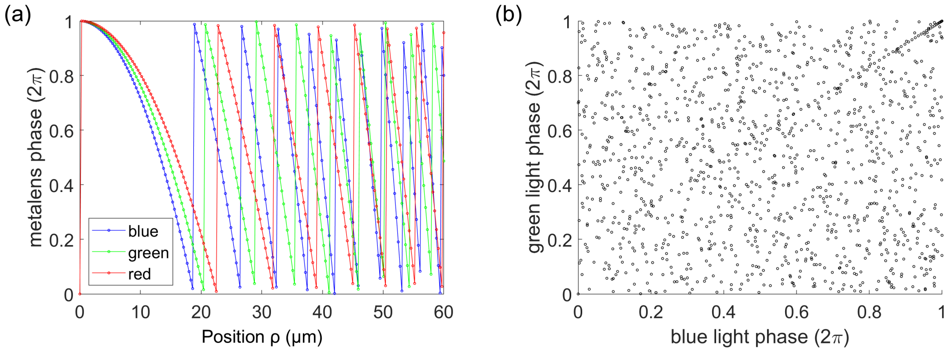

FIGURE 6.1: **(a)** The wrapped phase profile ($0 < \rho < 60\,\mu\text{m}$) of the polychromatic metalens working at red (643 nm), green (532 nm), and blue (444 nm) wavelengths, obtained from the ray-tracing optimization described in Chapter 4.3.5. **(b)** Required meta-atom phase modulations at green and blue wavelengths that perfectly match the phase profiles in (a).

and small apertures on the order of the wavelength. Such compact lenses are suitable for niche applications, such as direct integration with single-mode fibers [23], but are inadequate for most imaging endoscopes, which require larger apertures and high NA.

In large-aperture metalenses, the phase profile must be wrapped with 2π discontinuities, making achromatic correction more difficult. To eliminate chromatic aberration in this case, the phase wrapping points would need to shift with wavelength—a condition that cannot be satisfied using conventional dispersion engineering of meta-atoms. Although techniques like extended-depth-of-focus (EDOF) metalenses offer a workaround [24], they do so at the cost of reduced resolution and increased reliance on computational post-processing.

In the context of tri-color RGB scanning fiber endoscopes, where monochromatic laser light at red, green, and blue wavelengths is used for imaging, the requirement for continuous broadband achromatic performance can be relaxed. Instead, a polychromatic metalens—designed to exhibit identical optical functionality at several discrete wavelengths—can effectively meet system requirements. Since it only needs to satisfy the phase profiles at a finite set of target wavelengths, the design complexity is significantly reduced compared to that of the broadband achromatic metalenses π.

We design a polychromatic metalens for tri-color RGB scanning fiber endoscopy (SFE), which must closely satisfy three distinct wrapped phase profiles corresponding to red (643 nm), green (532 nm), and blue (444 nm) wavelengths. These phase profiles are obtained from ray-tracing simulation, as described in Chapter 4.3.5. They exhibit different 2π discontinuity points, as shown in Figure 6.1a, making direct meta-atom-to-phase mapping nontrivial. Achieving perfect simultaneous mapping requires the meta-atoms to support phase modulations denoted as $(\phi_\text{R}, \phi_\text{G}, \phi_\text{B})$ that span the entire $(0, 2\pi) \times (0, 2\pi) \times (0, 2\pi)$ volume [25, 17]. Figure 6.1b illustrates the distribution of the required $(\phi_\text{G}, \phi_\text{B})$ phase pairs as an example, which fully covers the entire $(0, 2\pi) \times (0, 2\pi)$ range.

To meet the phase requirements at all three wavelengths as much as possible, we employ a cross-shaped meta-atom. Compared to simple geometries such as squares or cylinders, the cross shape provides a better phase coverage in the $(0, 2\pi) \times (0, 2\pi) \times$

$(0, 2\pi)$ phase space, as supported by prior studies [26]. At the same time, it remains sufficiently simple to enable reliable and reproducible fabrication, unlike more complex structures.

6.3 Inversed design of a polychromatic metalens

We credit Zhihao Zhou for performing the inversed design of the polychromatic metalens.

6.3.1 Design framework

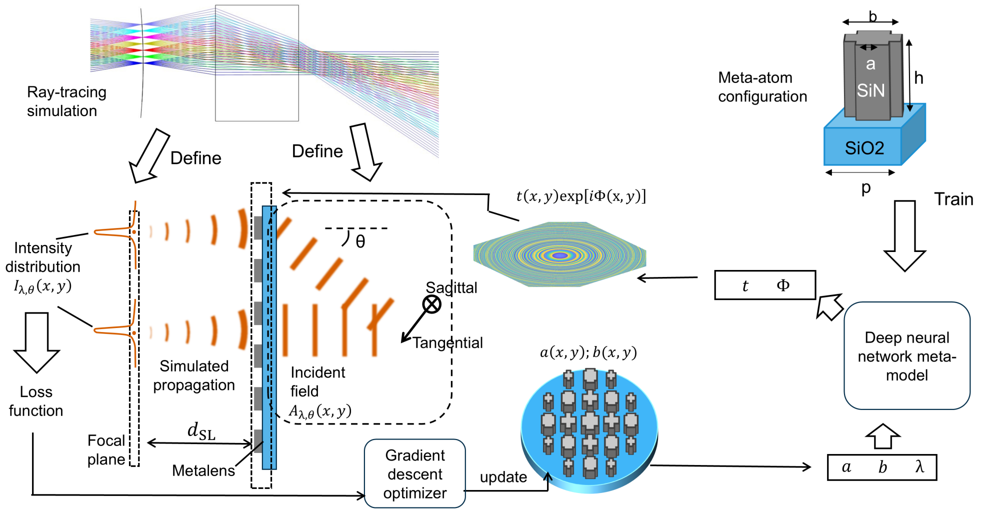

FIGURE 6.2: Inverse design approach of the RGB polychromatic metalens as the projecting lens in the SFE. The positions and incident angles θ of the incident beams are obtained by reversing the propagation direction of the projected beams in the ray-tracing simulation, the meta-atoms are Si_3N_4 on SiO_2 cross-shape structures with two tunable parameters a and b, pitch $p = 300\,\text{nm}$, height $h = 750\,\text{nm}$. The working wavelengths λ are 444 nm, 532 nm, and 643 nm. The output of the meta-model t and ϕ are the transmission and phase modulation of the meta-atoms.

Using the cross-shaped meta-atoms, a trivial approach for designing the metalens is to select a meta-atom at each lattice site of the metalens by minimizing the phase error between the desired and the achievable phase profiles at the three target wavelengths. However, this local forward optimization does not necessarily lead to optimal global performance, as the overall functionality of the metalens is determined collectively by the phase and amplitude contributions of all meta-atoms. To address this limitation, we implement an inverse design framework, illustrated in Figure 6.2, to iteratively optimize the geometric parameters of all meta-atoms across the metalens.

The inverse design is performed via minimizing a loss function that evaluates the metalens performance using a gradient descent-based optimization algorithm. For convenience in defining the loss function, we reverse the beam propagation direction compared to the actual SFE imaging head operation—treating the projected beams as incident beams—justified by the system's reciprocity.

Thus, the functionality of this metalens for SFE becomes:

Gaussian beams with a diameter d_{bL} incident from various angles θ and at the three design wavelength λ_i on the metalens at the location y, which is described by a complex scalar field $A_{\lambda_i\theta}(x,y)$. The incident angles and positions of the beams correspond to those of the projected beams in the SFE imaging head, determined via ray-tracing simulations (Chapter 4.3.5).

The metalens impart a transmission and phase modulation, $t(x,y)\exp(\phi(x,y))$. This modulation is calculated via a meta-model from the metalens designs, $a(x,y)$ and $b(x,y)$, where (a,b) are the geometric parameters of the meta-atoms, and (x,y) are the meta-atom lattice sites on the metalens.

The output field from the metalens then propagates to the focal plane where the fiber tip is physically scanned, which is at a distance d_{SL} away from the metalens. This propagation is simulated via the angular spectrum method, which is discussed in Chapter 2.2.2, to obtain the light intensity distribution on the focal plane, $I_{\lambda,\theta}(x,y)$.

The goal is finding a design $a(x,y), b(x,y)$ to get all the incident beams well focused. Starting from a random initialization of geometry, we apply a fully differentiable gradient descent method [26] to optimize the values of $a(x,y), b(x,y)$ by minimizing the loss function, which is computed from the intensity profile on the focal plane to evaluate how well the incident beams are focused. Additional details of the inverse design framework can be found in [26].

6.3.2 meta-atom model

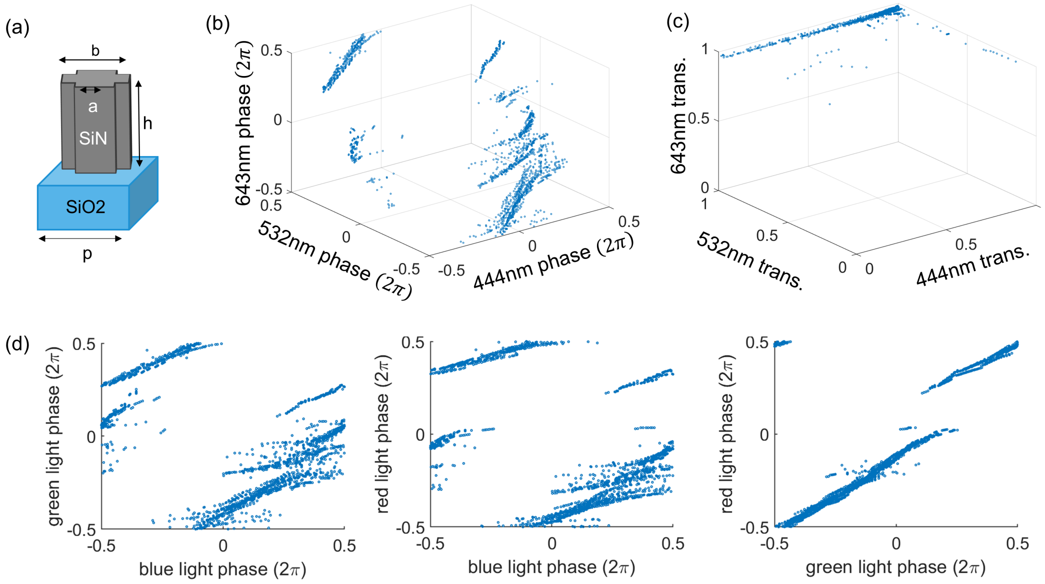

FIGURE 6.3: **(a)** meta-atom unit cell configuration. **(b,c)** The phase modulation (b) and transmission (c) distribution of meta-atoms with $80\,\mathrm{nm} \le a \le b \le 260\,\mathrm{nm}$ at the three design wavelengths obtained via FDTD simulation. **(d)** The phase modulation of the meta-atom at two of the three design wavelengths, plotted in 2D diagram.

To perform the fully differentiable inverse design framework to iteratively optimize the geometric parameters of all the meta-atoms across the entire metalens [26], differentiable functions of the meta-atom's phase and amplitude modulations on the incident light with regard to its geometric parameters, $t(a,b)$ and $\phi(a,b)$, are essential.

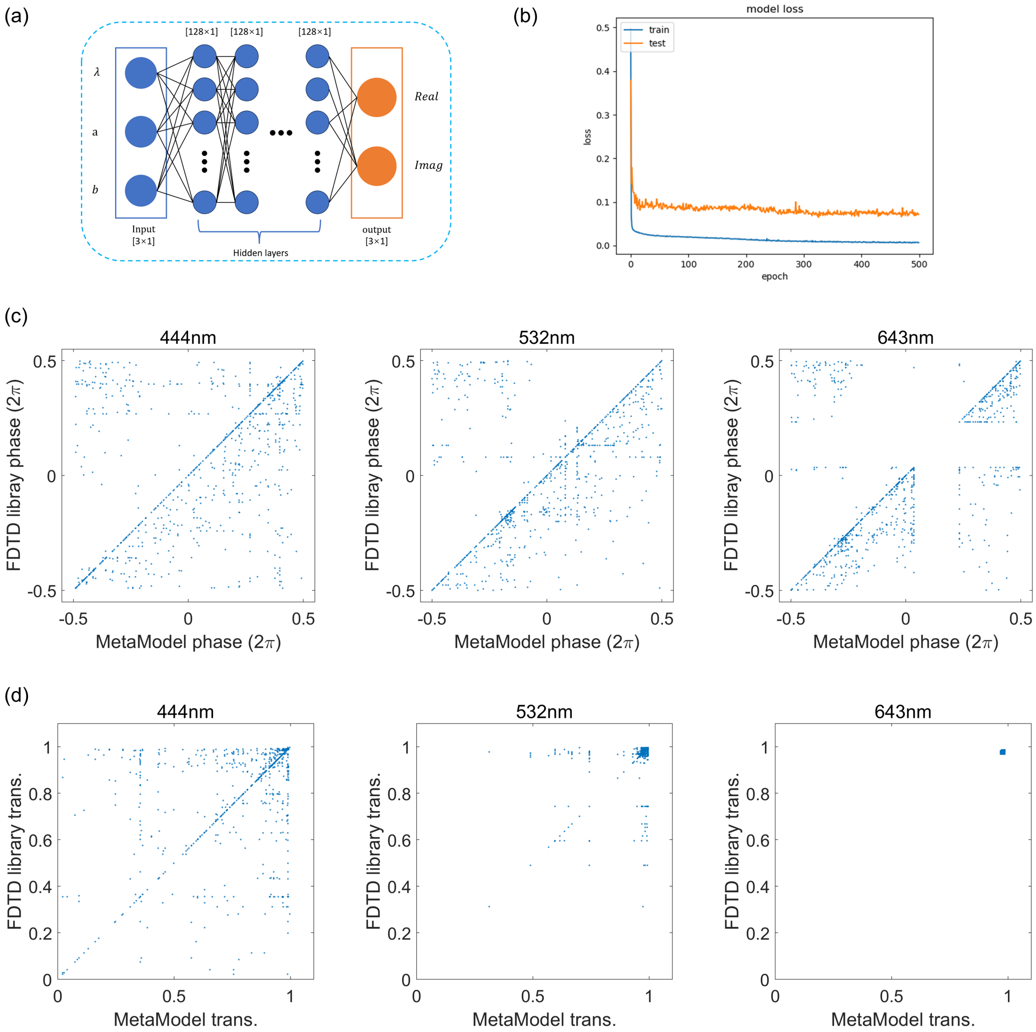

FIGURE 6.4: **(a)** The architecture of the DNN meta-model. **(b)** The loss function evaluation during the training process of the DNN. **(c,d)** The phase retardation (c) and amplitude modulation (d) of the meta-atom, DNN predicted values vs. FDTD simulated values (regarded as ground truth) in the validation process of the DNN.

These functions are approximated by a deep neural network (DNN) meta-model, because the DNN model has a better fitting capability to address drastic fluctuations of the phase and amplitude modulations of those meta-atoms that are in resonance modes.

To train this DNN meta-model, 5673 training samples of the meta-atoms' phase and amplitude modulations are calculated as meta-atom library via finite-difference time-domain simulations with monochromatic plane wave incident and periodic boundary conditions. Figure 6.3a shows the meta-atom configuration. The height of the Si_3N_4-on-SiO_2 meta-atoms is 750 nm. The unit cell is square with a periodicity of 300 nm. The two tunable dimension parameters a and b range from 80 nm to 260 nm, and satisfy $80\,\text{nm} \leq a \leq b \leq 260\,\text{nm}$. The wavelengths are 643 nm, 532 nm, and 444 nm. The training dataset is generated via finite-difference time-domain simulation. Figure 6.4b-d plots the phase and transmission distribution of these meta-atoms.

As can be seen in Figure 6.4a, the DNN architecture consists of a 10-layer fully connected network (FCN), with 128 units per layer and a Rectified Linear Unit (ReLU) activation function. The input features of the DNN comprised the incident wavelengths and the two geometric parameters of the meta-atom, while the output consisted of the real and imaginary components of the complex transmission of the meta-atom. In the training process of the DNN meta-model, the dataset is the meta-atom library displayed above, with 95% allocated for training and 5% for validation. The DNN networks were optimized by minimizing the mean squared error (MSE) between the predicted values and the ground truth (the loss function for the DNN training). This optimization was performed using the Adam optimizer with a learning rate of 10^{-4}, and the model was implemented within the TensorFlow framework. The evolution of the loss function during the DNN training process is shown in Figure 6.4b. The computational hardware consisted of a six-core CPU operating at 2.40 GHz, 23 Gigabytes of RAM, and one NVIDIA Tesla V100 GPU. The training process required 13 minutes for 500 iterations, while prediction generation during validation took 27 milliseconds. The validation results of the trained DNN meta-model are plotted in Figure 6.4c and d.

6.3.3 Loss function

The loss function is calculated from the intensity profile on the focal plane, which is defined as:

$$L = -\sum_{\lambda,\theta} \log(\mathrm{SR}_{\lambda,\theta}) \tag{6.1}$$

where the summation is performed over seven beam projecting angles θ as in the ray-tracing simulation in Chapter 4.3.5, and three wavelengths λ. Here $\mathrm{SR}_{\lambda,\theta}$ is the Strehl ratio of the incident beams at θ and λ. The Strehl ratio that defines the focusing efficiency of the lens is calculated as the maximal value of the simulated light intensity distribution $I_{\lambda,\theta}(x,y)$ on the focal plane divided by the peak intensity on the focal plane of an ideal lens. We empirically define this loss function because the percentage of the collimated incident light being focused to a spot is proportional to the Strehl ratio. In addition, we have tried different loss functions, and found that minimizing this negative summation of the log of the Strehl ratio will uniformly maximize the Strehl ratio at all incident angles and wavelengths.

6.4 Simulated performance of the metalens

We use the angular spectrum method discussed in Chapter 2.2.2 to simulate the performance of the designed polychromatic metalens. First we analyze the beam intensity distribution in the optical (y-z) plane (longitudinal distribution). The beam emitted from the fiber tip has a Gaussian distribution with a divergence angle defined by the numerical aperture (NA) of the fiber, which is set to 0.1 to match the NA of the single mode fiber we use in later experiments:

$$\tan\phi = \frac{\lambda}{\pi w_0} = \mathrm{NA} = 0.1 \tag{6.2}$$

where ϕ is the divergence angle of the Gaussian beam and w_0 is the waist width.

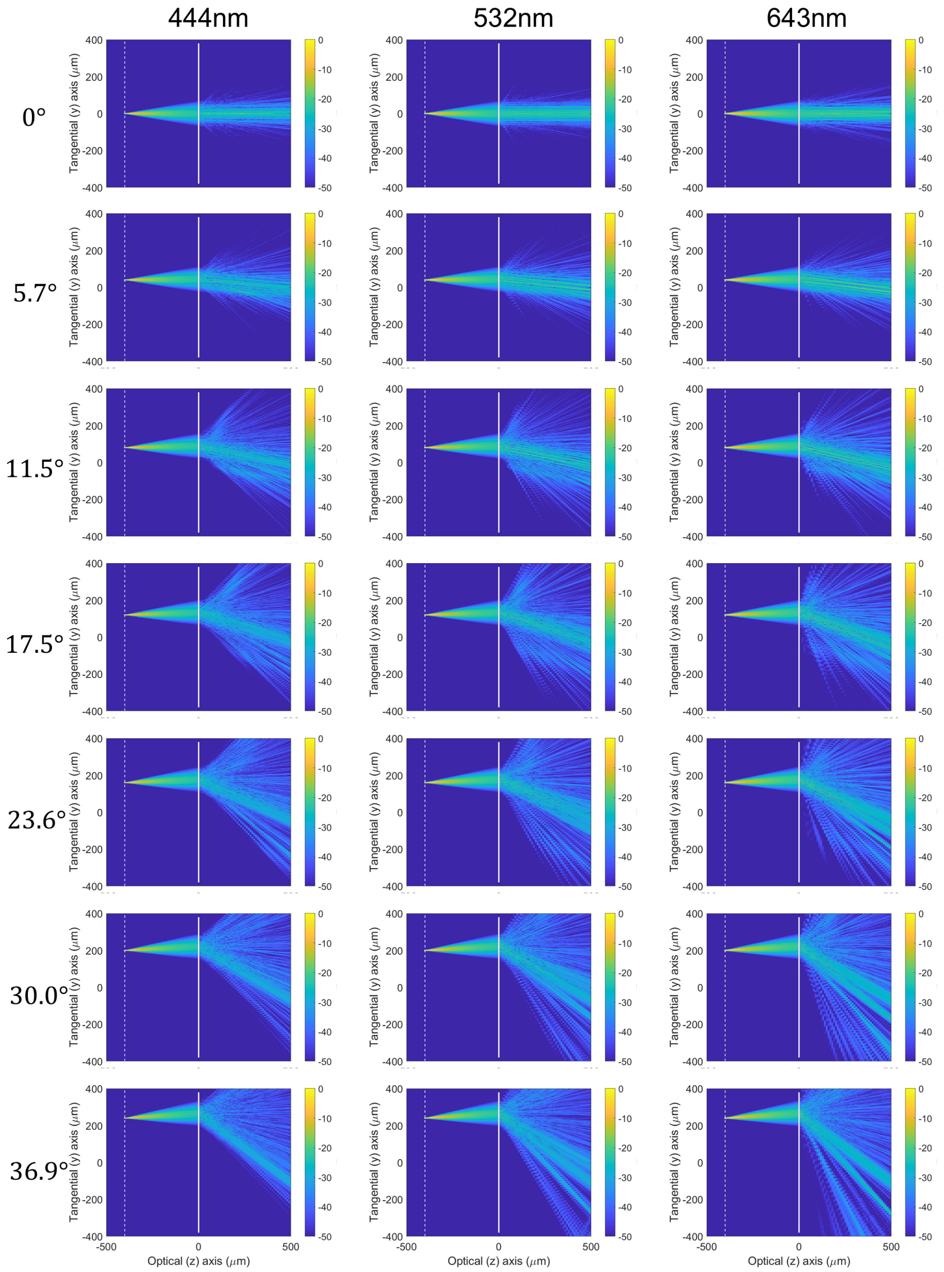

FIGURE 6.5: Longitudinal (y-z) plane intensity distribution of the beams being projected by the metalens. The beams emit from fiber tips at 7 different positions on the focal plane (dashed white line), and projected to 7 different angles, following the relationship in Figure 4.3d.

The intensity distributions for different wavelengths and projecting angles in the optical plane are summarized in Figure 6.5. Notably, the beams after transmitting the metalens are collimated and projected to the same θ for different wavelengths.

We further analyzed the angular intensity distribution in the transverse plane (i.e., perpendicular to the propagation direction of the beam) as plotted in Figure 6.6a and b, which essentially determines the imaging quality of the SFE. These angular distributions are calculated using the intensity distributions in k-space, which is obtained by performing a two-dimensional Fourier transform of the complex field of the beam after transmitting through the metalens. As can be seen, the projected beams have a slightly diverging angle due to their finite spatial extent, resulting in

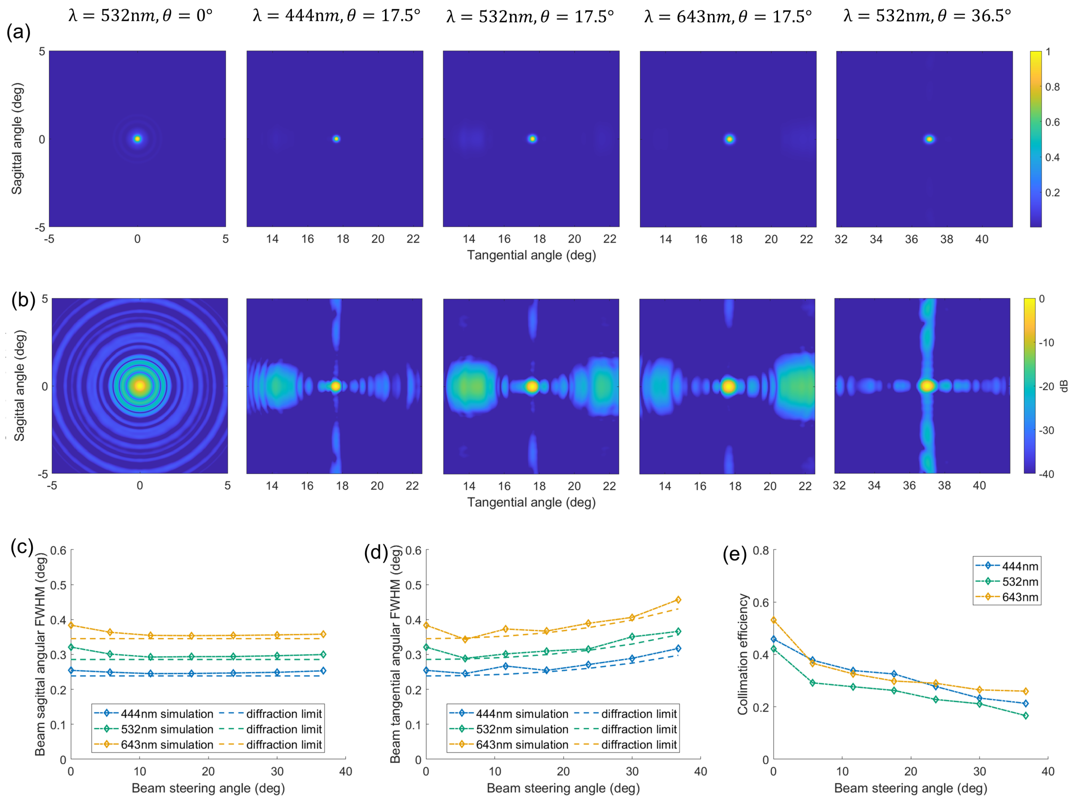

FIGURE 6.6: **(a),(b)** Simulated angular intensity distribution of the projected beams for different λ and θ at the far field. These distributions are normalized by the maximum intensity, plotted on a linear scale in (a) and on log scale in (b). **(c),(d)** The simulated and diffraction limited angular diameter of the projected beams in sagittal (c) and tangential (d) direction at the far field as a function of the beam projecting angle. For calculating the diffraction limits, the projected beam is assumed to be a Gaussian beam whose waist position is at the metalens and the amplitude profile at the waist equals the incident beam on the metalens. **(e)** The relative collimating efficiency as a function of the beam projecting angle. The data is calculated from the numerical simulation.

angular spreads of energy at the far field (the Fresnel factor $F = d_{\mathrm{bL}}^2/4\lambda R_{\mathrm{im}} \ll 1$, where $d_{\mathrm{bL}} = 2d_{\mathrm{SL}} \cdot \mathrm{NA}$ is the diameter of the incident beam on the metalens, λ is the wavelength, and R_{im} is the propagation distance of the projected beam). However, the beam is mostly collimated, evident by the high intensity in a confined angular section (normalized intensity > 0.1).

To better characterize the system, we define two quantitative figures-of-merit: the angular diameter of the projected beams and the relative beam collimating efficiency. The beam angular tangential (sagittal) diameter is defined as the full-width-at-half-maximum (FWHM) of the one-dimensional cross-section of the angular intensity distribution along the tangential (sagittal) axis. This diameter effectively gauges the angular resolution of the SFE image. We calculated the projected beams' angular diameters from the simulated angular intensity distribution, and compared them with the angular diffraction limits (discussed in Chapter 4.3.4, calculated using the Equation 4.20).

As can be seen in Figure 6.6c and d, the calculated angular diameter as a function of projecting angle matches closely with the diffraction limit. These beam angular diameters are proportional to the wavelengths, and the tangential diameters are inversely proportional to $\cos\theta$. These results are as expected from the angular resolutions of a typical optical imaging system. The projected beam has a diffraction limited angular diameter of under $0.45^\circ, 0.38^\circ, 0.32^\circ$ for red, green, and blue wavelengths. This could be improved by increasing the NA of the fiber and thus the effective aperture.

We also note that in the simulated angular intensity distributions of the projected beams in Figure 6.6d, there exist sidelobe spots with lower intensity but larger sizes compared to the collimated beam spot. Only a certain part of the total energy is concentrated within the collimated beam spot. Thus we define the relative collimating efficiency of our RGB metalens, which is the energy concentrated within the diffraction limited angular diameter of the projected beam divided by the total energy transmitting through the metalens. This efficiency will affect the imaging contrast in SFE. Figure 6.6e plots the simulated relative collimating efficiency of the designed metalens for 3 design wavelengths at different beam projecting angles. In simulation, 3 design wavelengths show similar efficiencies of around 50% at 0° projecting angles, and the efficiencies decrease with increasing projecting angles, down to around 25% at the maximal design projecting angle of 35°.

6.5 Metalens fabrication

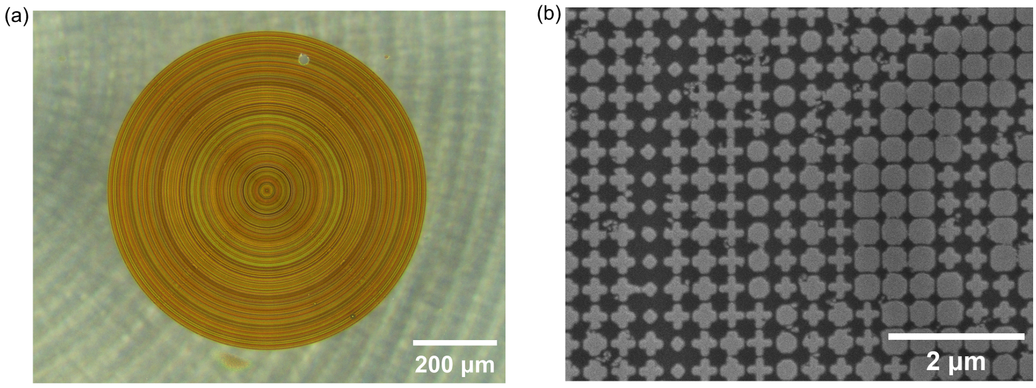

FIGURE 6.7: **(a)** Top view optical microscope and **(b)** top-down view SEM images of the fabricated RGB metalens.

The metalens is fabricated using an electron beam lithography process similar to the near infrared lens in Chapter 5. The only difference are the materials of the metalens layer and the substrate. A 750 nm thick Si_3N_4 is deposited on a 500 µm thick SiO_2 substrate using plasma enhanced chemical vapor deposition (SPTS Technologies Ltd., Delta LPX). An e-beam resist (ZEP-520A) is then spin-coated onto the chip at 5000 rpm and patterned by e-beam lithography (JEOL Ltd., JBX-6300FS). Subsequently, a $\sim$ 85 nm thick Al_2O_3 hard mask is created by electron beam assisted evaporation (CHA Industries, SEC-600) followed by resist lift-off in N-methyl-2-pyrrolidone (NMP) solution at 90°C overnight. Subsequently, the Si_3N_4 layer is etched by a fluorine based reactive ion process (Oxford, PlasmaLab 100, ICP-180). Figures 6.7a and b show the optical and scanning electron microscope images of the fabricated metalens.

6.6 Characterization of the metalens

6.6.1 Experimental setup

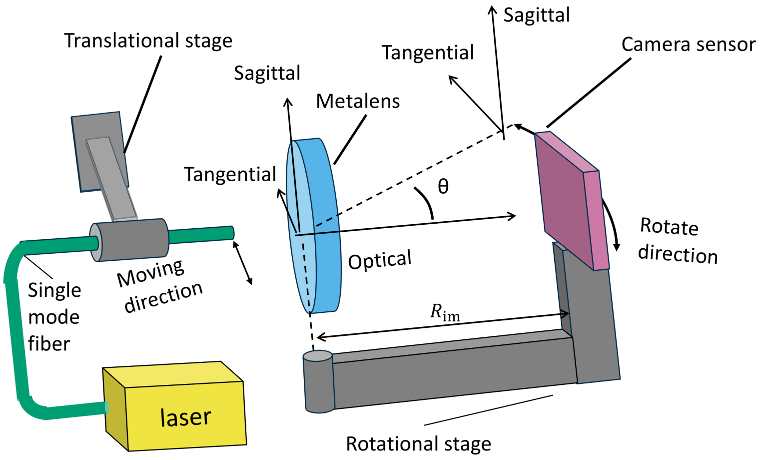

FIGURE 6.8: Experimental measurement setup to determine the angular intensity distribution of the collimated beam at various projecting angles.

To experimentally measure the angular intensity distribution of the projected beam at various projection angles, the setup illustrated in Figure 6.8 was constructed. A single-mode fiber (Thorlabs, P1-460B-FC-5) with numerical aperture NA $= 0.1$ was mounted on a 3D translational stage, with its tip adjustable within the focal plane of the metalens. The fiber was coupled to laser sources: a supercontinuum laser (NKT Photonics, SuperK Fianium FIR20) with a tunable filter (SuperK SELECT 4x VIS/IR) provided monochromatic light in the range of 485 nm to 643 nm, and a blue diode laser (Opto Engine LLC, MDL-III-445L 80,mW) supplied light at 444 nm. A CMOS camera sensor (Allied Vision, Prosilica GT 1930C) was mounted on a rotational arm, with the rotation axis co-axially aligned to the center of the metalens.

The fiber tip was first aligned with the metalens, using the same method described in Chapter 5.4.1. Then, as the fiber tip was translated parallel to the tangential axis, the emitted beam was projected by the metalens, emulating the scanning fiber tip in an SFE imaging head. The rotational arm was then adjusted along the tangential axis to capture the projected beams at different projecting angles θ. The sensor-to-metalens distance was fixed at $R_{\mathrm{im}} = 80\,\mathrm{mm}$ to ensure far-field conditions, with a Fresnel factor of $F = a^2/R_{\mathrm{im}} = 0.18$, where $a = 2f \times \mathrm{NA} = 80\,\mu\mathrm{m}$ is the diameter of the incident beam on the metalens, and $\lambda = 440\,\mathrm{nm}$ is the shortest operating wavelength.

Figure 6.6b shows the sidelobe spots with an intensity much lower than that of the collimated beam. To experimentally capture these sidelobe beams, it is required to resolve intensity variations spanning five orders of magnitude. This is achieved by capturing multiple images at different exposure times, ranging from approximately 100 µs to 1 s. Saturated pixels from longer exposures were subsequently replaced with corresponding unsaturated pixels from shorter exposures to generate high dynamic range (HDR) composite images.

To quantitatively analyze the beam distribution across a large angular range, the one-dimensional cross-section of the angular intensity distribution for the on-axis beam (projecting angle $\theta = 0$) over a 0° to 30° angle range was subtracted from the two-dimensional angular intensity distribution measured across a $-2.5°$ to $2.5°$ sagittal angle and a $-2.5°$ to $32.5°$ tangential angle range. Due to the limited field of view of the camera sensor, each image captured only a $5° \times 10°$ angular segment. To obtain a full 30° tangential angular field, seven images were stitched together, each centered at tangential angles from 0° to 30° in 5° increments.

6.6.2 Angular intensity distribution

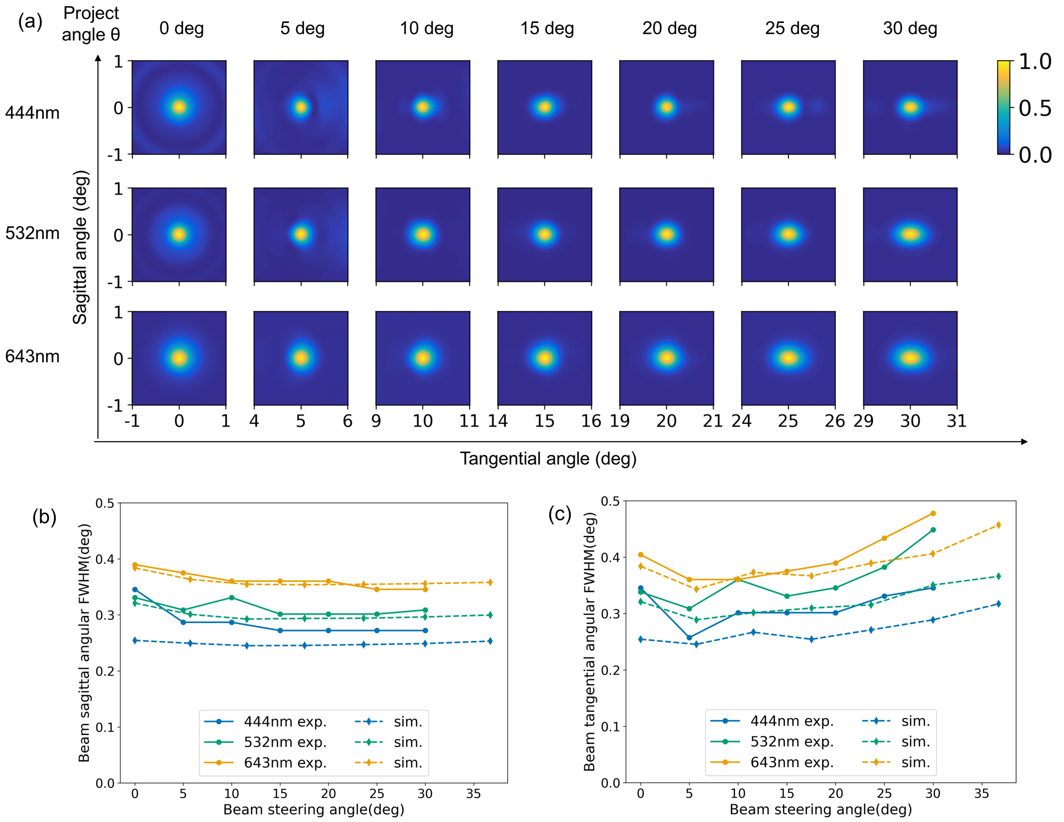

FIGURE 6.9: **(a)** Angular intensity distribution of the beam projected to $\theta = 0°$ - $30°$ at the three designed wavelengths. **(b)(c)** The experiment, simulation, and diffraction limited angular intensity FWHM of the projected beams in sagittal (b) and tangential (c) direction vs. the beam projecting angle for three design wavelengths.

Figure 6.9a shows the experimentally measured angular intensity distribution for beams projected from 0° to 30° for the three different wavelengths on a linear scale. Importantly, the beams are collimated and projected without chromatic dispersion to the same projecting angle θ, matching simulations.

We further calculated the experimental beam diameters from these angular distributions and presented the results in Figure 6.9b and c. Across the required projecting range, the experimental values are close to the simulation values and the diffraction limits, thus closely matching the desired functionality.

6.6.3 Collimating efficiency

To better illustrate the losses of the beam energy that lower the collimating efficiency, we also plotted the angular beam intensity distribution on a log scale in Figure 6.10a.

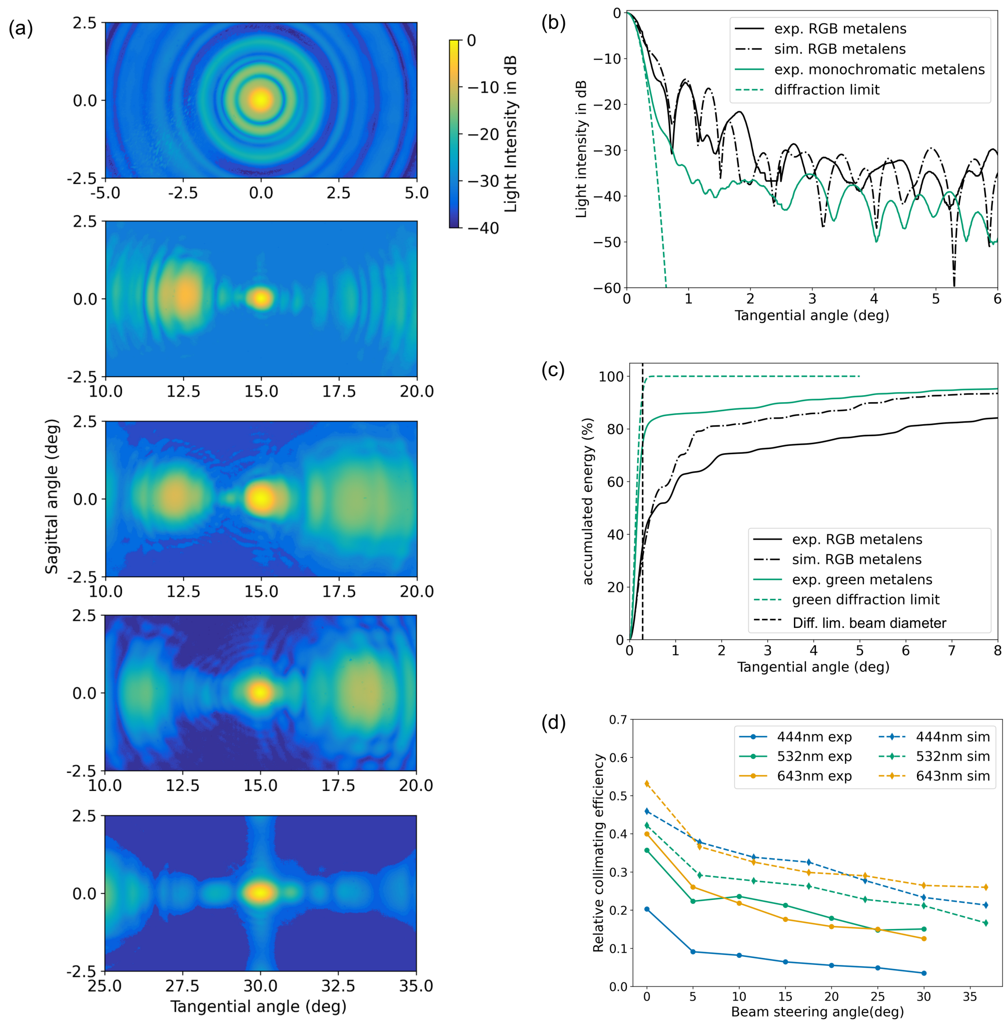

FIGURE 6.10: **(a)** The one-dimensional cross-sectional intensity distribution vs. tangential angle. **(b)** Accumulated energy in percentage within a certain range of angle from the center of the beam. For (a) and (b), the steered beam is at green (532 nm) and is steered to $\theta = 0$. The black solid (dashed) curve is the experimental (simulation) data from the RGB polychromatic metalens, while the blue solid (dashed) curve is the experimental (theoretical limit) data from a monochromatic metalens that has the same optical functionality but is designed to only work at green (532 nm). **(c)** Experimentally measured and simulated relative collimating efficiency vs. beam steering angle.

The measured angular intensity distributions match well with the simulated ones in Figure 6.6b, whereas losses primarily occur in sidelobes along the tangential direction with about a factor of 10 or much less than the intensity in the main lobe.

We calculated the on-axis $\theta = 0$ collimating efficiency by measuring the one-dimensional (1D) intensity cross-section along the tangential angular axis for the on-axis beam. This distribution is plotted in Figure 6.10b. It can be seen that at angles smaller than 0.3° (the diffraction limited projected beam's angular diameter), the 1D angular intensity distribution of the polychromatic RGB metalens overlaps with the diffraction limited distribution. However, at larger angles, the distribution of the RGB metalens drops slower than that of the diffraction limit. This indicates that the RGB metalens collimates the beam close to the diffraction limit, but distributes more energy outside of the FWHM of the beam, essentially leading to a lower efficiency. Figure 6.10c plots the accumulated energy vs. the angle by integrating the intensity vs. angle data in Figure 6.10b. As can be seen, for the RGB metalens only around 40% of the energy is concentrated within the diffraction limited angular diameter.

We obtained the experimental relative collimating efficiency at $\theta = 0$ using the approach presented above. For $\theta \neq 0$, whose 2D angular intensity distributions are not central symmetric, the relative efficiency is calculated by measuring the absolute intensity of the projected beam at $\theta \neq 0$ and comparing with that of the projected beam at $\theta = 0$. This is valid, because the amount of light transmitted through the metalens does not vary with the projecting angle θ, which is supported by the simulation. Figure 6.10d shows the relative collimating efficiency of the RGB metalens vs. the beam projecting angle. Our polychromatic metalens has an average relative efficiency of around 32% for on-axis (0° projecting angle) among the 3 wavelengths, and an overall average efficiency of 13% across the entire 70° field-of-view (weighted by the solid angle). Specifically, red and green wavelengths show similar efficiencies around 40% for on-axis angle, which drops to around 15% for 30° projecting angle. These efficiencies are only about $70 - 80\%$ of the simulation results, which could be due to imperfections in fabrication and optimization model predicting the phase modulation and transmission of the meta-atoms, as well as inaccuracies in the beam propagation simulation based on the scalar-field diffraction theory. We note that unlike in the simulation results, the blue wavelength 444 nm has a notably lower experimental efficiency compared to red and green wavelengths, ranging from only 20% at 0° to 5% at 30°. This might be due to the fact that at this wavelength, the ratio of wavelength to the 300 nm pitch of the meta-atom is lower, resulting in a lower sampling rate of the light field by the metalens.

6.7 SFE Imaging test using the metalens

The imaging performance of our metalens was tested on an SFE platform invented in a previous work [10] and built by VerAvanti Inc.

We verified the imaging performance of our RGB metalens in an SFE testing platform, shown in Figure 6.11a. Three different laser outputs are coupled into the same single mode fiber, which is actuated by a piezo tube (PZT) actuator. The beam emitted at the fiber tip is projected by the projecting lens to illuminate a checkerboard test pattern. The reflected light is collected by a detector to construct the image (details in Chapter 4.2). The RGB image of the pattern captured by the metalens is shown in Figure 6.11b.

We note that the image appears non-uniform due to the position of the sensor relative to the imaging target (for an actual implementation of the SFE, several collection fibers are used simultaneously to form a uniform image). Therefore, we cropped the image to the upper right quadrant and displayed the respective color channels in Figure 6.12a. For comparison, we show images captured by using 3 different monochromatic metalenses designed for the corresponding wavelengths (Figure 6.12c). These images are linearly rescaled to enhance the visual representations. In Figure 6.12b and d, we plot the greyscale value variation across the checkerboard pattern. The plot reveals that at the interface of the bright and dark squares, the relative gradient of pixel values for the RGB metalens closely matches that of the monochromatic lens, suggesting that the RGB metalens maintains a comparable imaging resolution. However, it is notable that for the RGB metalens, the contrast

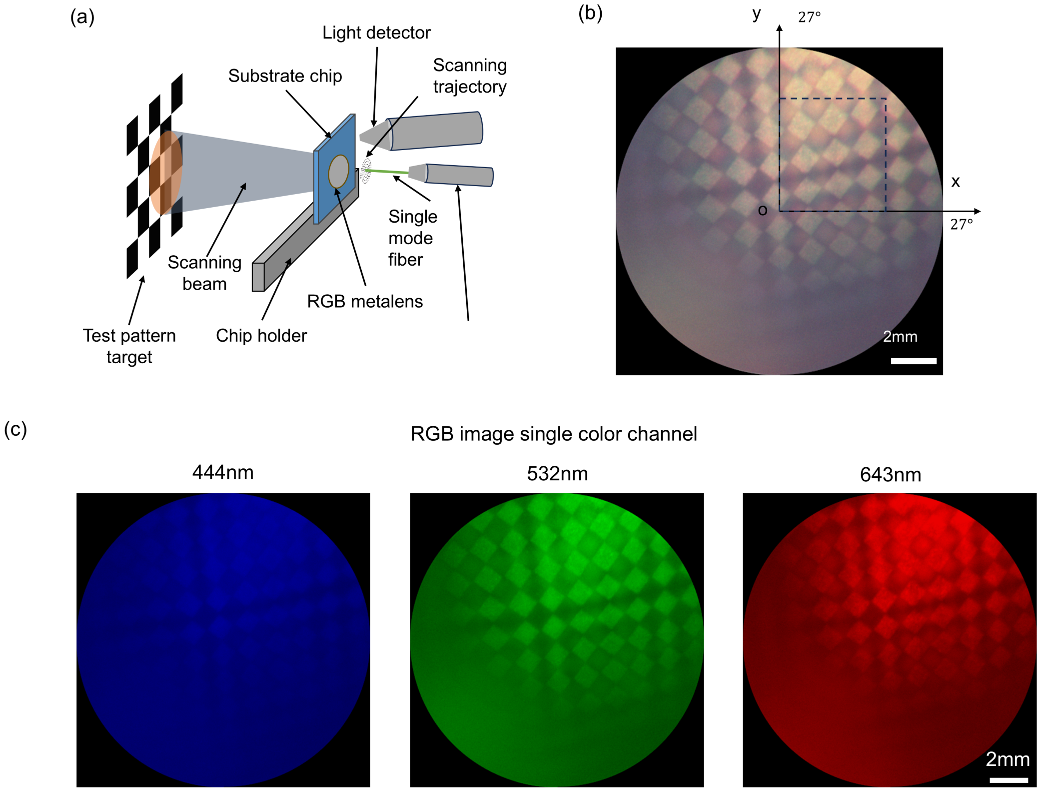

FIGURE 6.11: **(a)** Schematic demonstration of the scanning fiber endoscope imaging system. The imaging distance is 14.5 mm. The blocks of the checkerboard have a side length of 1.1 mm, and the total angular FOV of this SFE platform is 54°, limited by the actuation range of the single mode fiber. **(b)** Unprocessed RGB tri-color image of the test pattern using our polychromatic metalens. **(c)** Unprocessed image in (b) displayed in 3 color channels.

between the bright and dark blocks is less pronounced compared to the monochromatic lens. We define the pattern contrast as:

$$\text{Contrast} = \frac{I_{\text{bright}} - I_{\text{dark}}}{I_{\text{bright}} + I_{\text{dark}}} \tag{6.3}$$

where I_{bright} and I_{dark} are the average pixel values of the bright and dark block of the checkboard pattern. The patterns captured by the RGB metalens have a lower contrast by a factor of 4 for blue, 2.5 for green, and 2 times for red compared to that captured by the corresponding monochromatic metalens, owing to the low relative efficiency of the RGB metalens.

As can be seen in the images from the polychromatic RGB metalens, a blurry shadow of the checkerboard patterns overlays with the clear pattern. This shadow defect is likely caused by spurious side lobes, as shown in Figure 6.10a. These artifacts could be mitigated by image deconvolution. Specifically, the Richardson–Lucy deconvolution algorithm is a promising solution as it can accommodate for the position-dependent point spread function of our metalens. Alternatively, these artifacts could be reduced using a confocal SFE setup [27], in which the scanning fiber not only emits the light for illumination, but also collects the back-scattered light that is refocused by the metalens into the fiber. The confocal setup can spatially filter out the light not being collimated at the designed position, which greatly reduces the

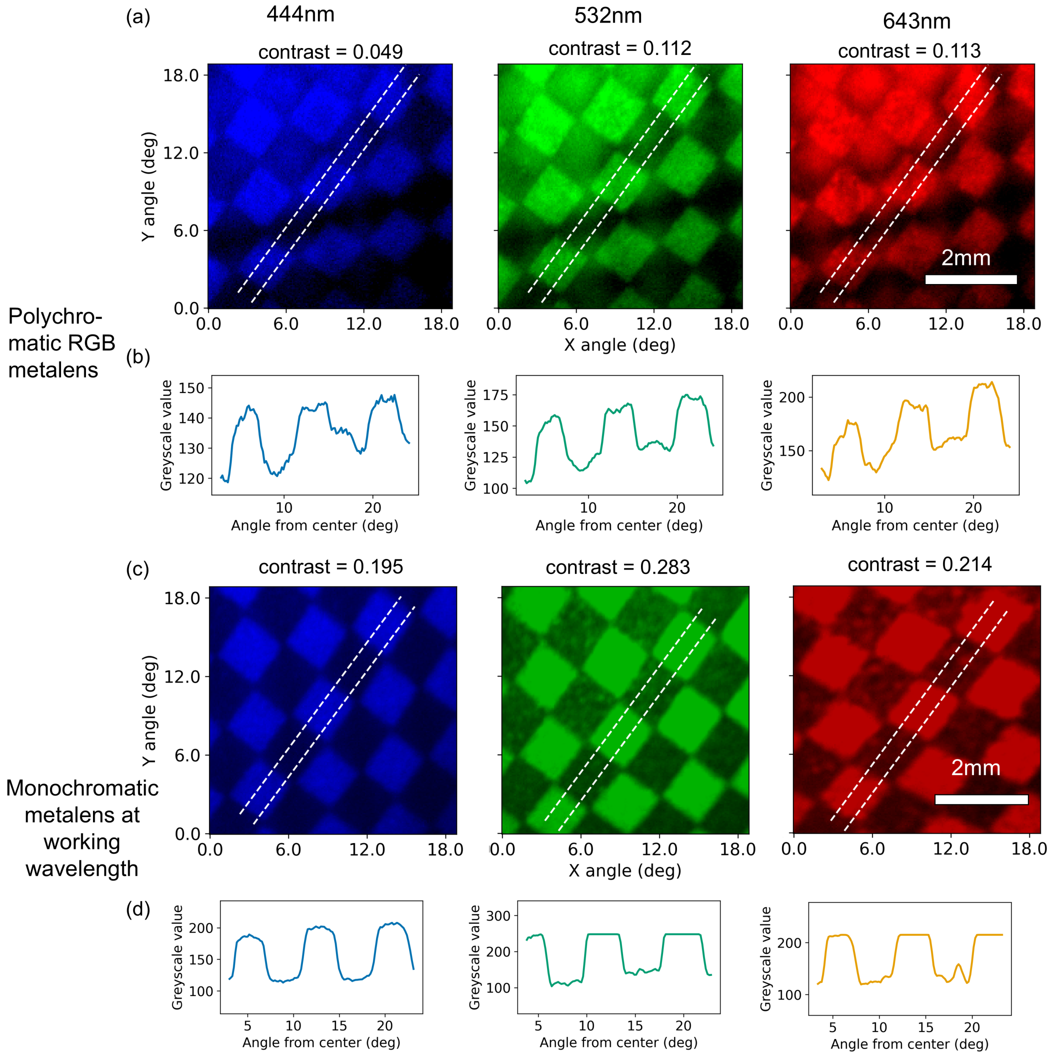

FIGURE 6.12: **(a)c)** (a) shows the blue, green, and red channel of the image in (b) within the zone circled by the dashed line square. (c) plots the same pattern imaged by 3 monochromatic metalenses designed for the corresponding wavelengths. A linear mapping of the pixel values of the raw images to that of these displayed images is tailored to each specific image to enhance the visual representations. The pattern contrast is defined as the difference divided by the sum of the average pixel values of the bright and the dark blocks. **(b)(d)** The greyscale value variations across the checkboard patterns along the two white dashed lines. The values are averaged over the region sandwiched by the two dashed lines.

effect of the spurious side beams. Specifically, in conjunction with a double-clad fiber [28] with a single mode core for emission and a larger multimodal clad for collecting return light, the lower collection efficiencies of a confocal setup which would otherwise impede such a device could be mitigated.

Chapter 7

Spectrally encoding imaging through a fiber

7.1 Overview

A spatial-to-spectral encoder (SSE) maps spatial intensity information (an image) onto the spectrum, which can be transferred without distortion through an optical fiber, followed by subsequent measurement of the spectrum and computational reconstruction of the image from spectral measurements. This approach allows for the transmission of multi-pixel image information through a single fiber core. This innovative strategy overcomes resolution limitations imposed by pixel density in conventional endoscopes (as discussed in Chapter 3.2.3) without resorting to any scanning mechanisms. As a result, it has the potential to significantly improve resolution without increasing the rigid tip length of the endoscope.

As a proof of concept, an array of miniaturized metasurface–Fabry–Pérot cavities are designed as an SSE, comprising 4×4 spatial pixels. With this spectral encoder, we demonstrate the transmission of a 4×4 binary image through a single fiber core. Unlike the SSE reported in the previous studies [11, 12] which relies on random medium, this encoder is designable, and thus serves as a starting point for the development of spectrally encoded, non-scanning endoscopic imaging systems.

We credit Quentin Tanguy for the initial idea of the metasurface–Fabry–Pérot cavities as an SSE.

This chapter is adapted from a published paper [29].

7.2 Spectrally encoded imaging fundamentals

7.2.1 Spatial-to-spectral encoding

As demonstrated in Figure 7.1a, given a wavelength independent spatial information $I(x, y)$, to encode it into a spectrum $I(\lambda)$, mathematically, the encoding process via a spatial-to-spectral encoder (SSE) can be expressed as:

$$I(\lambda) = \int_x \int_y M(x, y, \lambda) I(x, y) dx dy \tag{7.1}$$

where $I(\lambda)$ is the spectrum of light encoded with spatial information, $I(x, y)$ is the lateral intensity distribution of the light that contains spatial information, and

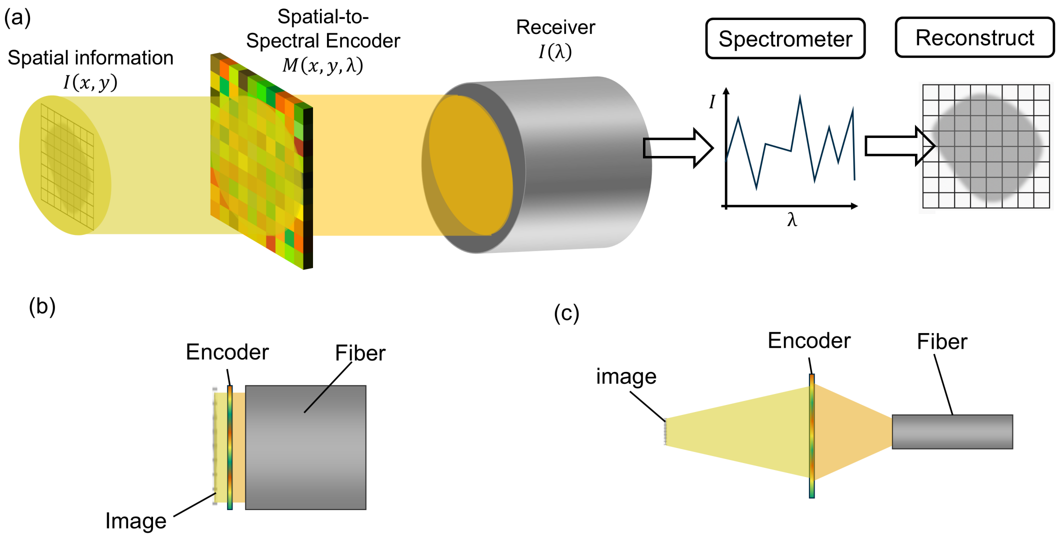

FIGURE 7.1: **(a)** Schematic demonstration of imaging through a fiber via spatial-to-spectral encoding and computational decoding. **(b)** Demonstration of near-field spectrally encoded imaging. **(c)** Demonstration of far-field spectrally encoded imaging.

$M(x, y, \lambda)$ is the spatial-spectral modulation of the SSE. The light incident from a lateral spatial position (x, y) transmits through the SSE with a transmission spectrum $t_{\mathrm{xy}}(\lambda) = M(x, y, \lambda)$. It can be written in a discretized and matrix multiplication format. The spatial position (x, y) is discretized into n spatial pixels, and the wavelength λ within the working spectral range is discretized into m levels.

$$\underbrace{\begin{bmatrix} I_{\lambda_1} \\ \vdots \\ I_{\lambda_m} \end{bmatrix}}_{a} = \underbrace{\begin{bmatrix} t_{11} & \dots & t_{1n} \\ \vdots & & \vdots \\ t_{m1} & \dots & t_{mn} \end{bmatrix}}_{\mathbf{M}} \cdot \underbrace{\begin{bmatrix} I_1 \\ \vdots \\ I_n \end{bmatrix}}_{b} \tag{7.2}$$

where λ_i is the discretized wavelength within the spectral range of the spectral encoder, m is the number of discretized levels, a is the discretized spectrum of the encoded light, and b is the vectorized spatial information with n pixels. Typically we have $m > n$. Each column in the matrix $\mathbf{M}$ is the discretized format of the transmission spectrum $t_{\mathrm{xy}}(\lambda)$,

$$(x, y) \to j \tag{7.3a}$$

$$t_{ij} = M(x, y, \lambda_i) \tag{7.3b}$$

7.2.2 Computational decoding

The encoded light with a spectrum $I(\lambda)$ can be transformed through a fiber and then be measured by a spectrometer. The measured spectrum can be computationally decoded to recover the spatial intensity information $I(x, y)$, as demonstrated in

Figure 7.1a. The decoding could be done via a pseudo inversion:

$$b = \mathbf{M}^{+} a \tag{7.4}$$

$$\mathbf{M}^{+} = (\mathbf{M}^{T}\mathbf{M})^{-1} M^{T} \tag{7.5}$$

The condition number of the encoding matrix **M**, $\kappa(\mathbf{M}) = \|\mathbf{M}\| \cdot \|\mathbf{M}^{+}\|$, is an indicator of the stability of the pseudo inversion process, which can be used as an FOM of the spectral encoder. Once the encoding matrix **M** is calibrated, the decoding process can be achieved by a computationally efficient matrix multiplication, which facilitates real-time imaging.

7.2.3 Near field and far field configuration

This imaging approach based on spectral encoding and decoding can be implemented in either the near-field (Figure 7.1b) or far-field (Figure 7.1c) configurations.

In the near-field configuration, both the input spatial intensity distribution $I(x, y)$ and the receiver of the spectrally encoded output are located in close proximity to the encoder, as illustrated in Figure 7.1b. Here, the encoder consists of an array of distinct spectral filters, where each filter at position (x, y) exhibits a unique normal-incidence transmission spectrum denoted by $t_{\mathrm{xy}}(\lambda)$.

In contrast, the far-field configuration involves both the input $I(x, y)$ and the receiver being positioned at a distance from the encoder, as shown in Figure 7.1c. In this case, the encoder functions as a lens with large and unconventional dispersion. Its point spread function (PSF) on the image plane, $\mathrm{PSF}_{\lambda} = M(x, y, \lambda)$, varies strongly with the wavelength λ. To achieve effective spectral encoding, the PSFs corresponding to different wavelengths must be highly orthogonal to one another—unlike in conventional metalenses, where chromatic dispersion primarily results in beam spots of different sizes due to varying degrees of defocus.

While the far-field encoder is well-suited for endoscopic systems employing dense fiber arrays, this chapter presents a proof-of-concept demonstration based on the near-field encoder configuration.

7.3 Design of SSE

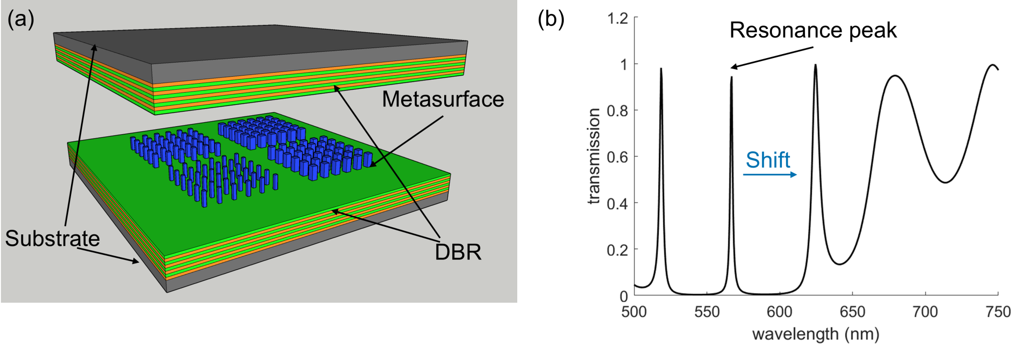

FIGURE 7.2: **(a)** Structures of a metasurface-Fabry-Perot cavity spatial-to-spectral encoder (SSE). **(b)** Working principle of the SSE in (a).

Spectral spatial encoders (SSEs) have previously been demonstrated using two orthogonally tilted Fabry-Pérot cavities [11] and random scattering media [12]. However, in these earlier implementations, the encoding matrix $M(x, y, \lambda)$ was inherently random and non-designable. This randomness resulted in device-specific spectral codes, necessitating individual calibration for each device and thereby limiting the scalability and practicality of spectrally encoded imaging systems. Furthermore, the random nature of the encoding matrix often led to a low condition number, rendering the decoding process—based on pseudo-inversion of the encoding matrix—highly sensitive to noise and calibration errors. Since the encoding matrix was not deliberately engineered, optimizing its performance was difficult. Additionally, the physical size of these SSEs was not compact, posing challenges for integration with miniaturized imaging devices such as endoscopes.

To address these limitations, we designed a near-field SSE, which utilizes an array of transmissive dielectric metasurface blocks embedded in a Fabry-Pérot cavity as phase-shifting elements, realizing a two-dimensional array of narrowband spectral filters with distinct passbands, as proposed in [30]. As illustrated in Figure 7.2a, the Fabry-Pérot cavity is formed by two opposite-facing distributed Bragg reflectors (DBRs), with an array of metasurface blocks embedded in the cavity. Each metasurface block introduces a phase delay that increases the effective optical path length, thereby red-shifting the resonant transmission peak of the cavity, as shown in Figure 7.2b. This spectral shift is proportional to the phase delay introduced by the metasurface blocks, which can be engineered by varying the lateral dimensions of its constituent meta-atoms.

We designed and fabricated a proof-of-concept metasurface-Fabry-Pérot cavity SSE with a 4 × 4 array of spatial pixels. Each pixel is an individual spectral filter that comprises a metasurface block with a distinct phase shift embedded within the cavity. The phase delay of each pixelated metasurface block, ϕ_j, effectively increases the round-trip phase delay ψ of the cavity by $2\phi_j$, thereby tuning the resonance wavelength $\lambda_{\text{res}-j}$ of each filter. To support a resonant mode within the DBR–metasurface –DBR cavity, the round-trip phase ψ must satisfy the condition:

$$\psi = 2\left(\phi_j + \frac{2\pi}{\lambda_{\text{res}-j}} L\right) = 2\pi q \tag{7.6}$$

where L is the cavity length and q is an integer corresponding to the order of the resonance, assuming normal-incidence plane-wave illumination. We set $L = 2.5\,\mu\text{m}$ and operate at the $q = 9$ mode. The resulting resonant wavelength is given by:

$$\lambda_{\text{res}-j} = \frac{2L}{q - \phi_j/\pi} \approx \frac{2L}{q} + \frac{2L}{q^2}\frac{\phi_j}{\pi} \tag{7.7}$$

The free spectral range (FSR) is the difference between adjacent resonance orders:

$$\Delta\lambda_{\text{FSR}} = \frac{2L}{q-1} - \frac{2L}{q} \approx \frac{2L}{q^2} \tag{7.8}$$

To ensure the encoding matrix has a low condition number, the transmission spectra of the spectral filters must be as orthogonal as possible. Thus, the resonance wavelengths $\lambda_{\text{res}-j}$ should be evenly distributed across the full FSR. This is achieved by setting ϕ_j to span from 0 to π in uniform steps. Meta-atoms with appropriate height, periodicity, and width are selected to realize the full 0–π phase range at the target wavelength range.

Each pixel in the 4×4 metasurface array measures 50 µm × 50 µm. The metasurfaces are composed of Si_3N_4 square meta-atoms on a fused silica substrate, with a height of 500 nm, a square lattice, and a pitch of 300 nm. For each j^{th} pixel, the meta-atom lateral size is chosen to produce a single-pass phase shift of $\frac{j}{16}\pi$ at the designed resonance wavelength. We employ rigorous coupled-wave analysis (RCWA) [15] to simulate the phase response as a function of lateral size and wavelength, generating a meta-atom library used for design selection.

The Fabry-Pérot cavity employs two 7-layer Si_3N_4–SiO_2 distributed Bragg reflectors as mirrors, providing the high reflectivity necessary for a narrowband resonance peak.

7.4 Fabrication of SSE

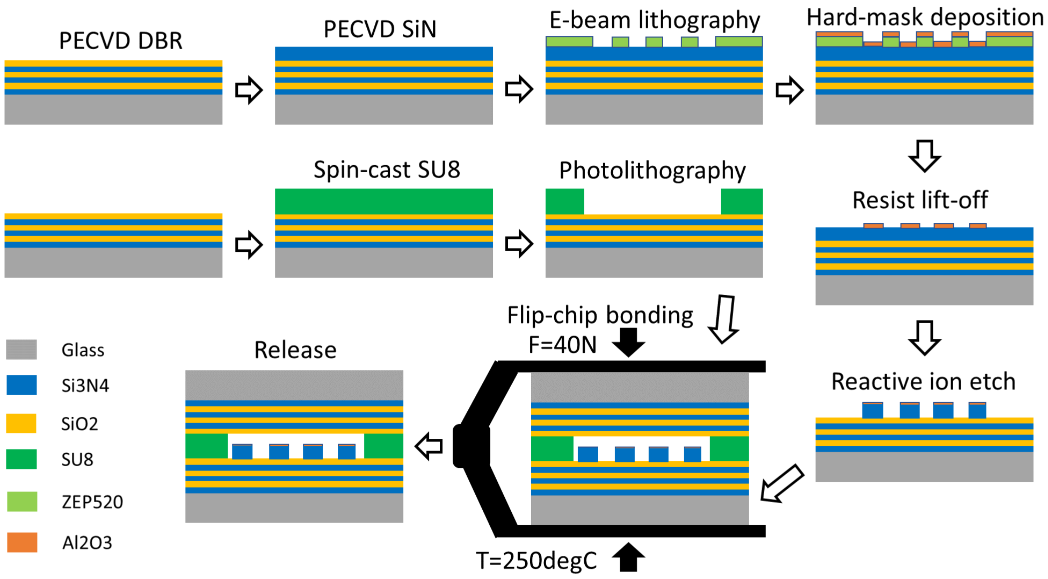

FIGURE 7.3: Fabrication process flow of the spatial-to-spectral encoder.

The two DBRs are grown on two 500 µm-thick fused-silica substrates ($10 \times 10\,\text{mm}^2$ chip) with plasma-enhanced chemical vapor deposition. These DBRs are made of 7 alternating layers of Si_3N_4 ($n = 1.97$ at 560 nm, thickness = 71 nm) and SiO_2 ($n = 1.44$ at 560 nm, thickness = 97 nm), which has a reflecting band centered at 560 nm. After that, 500 nm Si_3N_4 is deposited on one of the DBRs, where the metasurface is fabricated by electron beam lithography (EBL). A 300 nm-thick e-beam resist (ZEP-520A) is spin-coated and patterned by EBL, then a ≈ 65 nm thick Al_2O_3 hard mask is created by electron beam assisted evaporation followed by resist lift-off in N-methyl-2-pyrrolidone (NMP) at 90°C overnight. Subsequently, the Si_3N_4 layer is etched by a fluorine based reactive ion process. Meanwhile, a 2.5 µm-thick SU8 resist is spin-coated on the other DBR and patterned by photolithography to create a spacer structure (a ring with inner diameter = 5 mm, outer diameter = 8 mm, height = 2.5 µm).

Finally, the two DBRs are bonded together with the metasurface wrapped inside using a flip-chip bonder. The SU8 bonding is done by heating up the two substrates to 250°C and applying 40 N force for 1 hour.

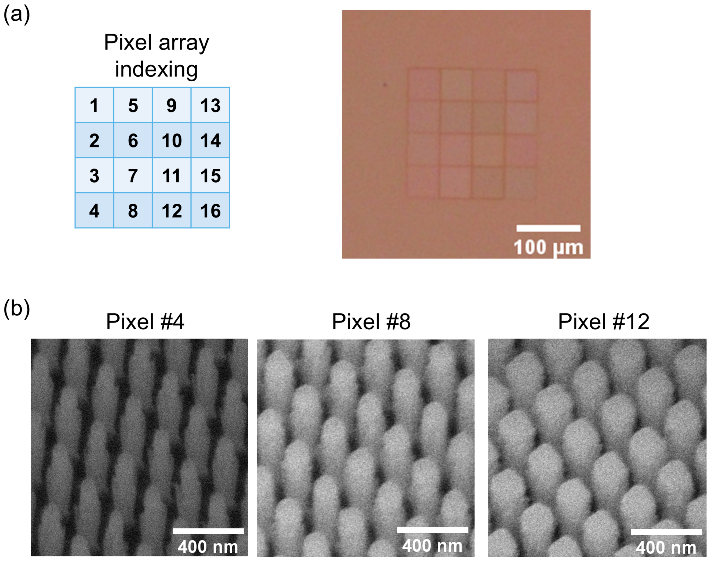

FIGURE 7.4: **(a)** Optical microscope top-down image of the fabricated spatial-to-spectral encoder. **(b)** Scanning fiber image of the metasurface layers of three of the pixels of the encoder. The images are captured in the middle of the fabrication process flow, before the two substrates are bonded.

Figure 7.4a shows an optical image of the fabricated SSE device from top, while Figure 7.4b shows scanning electron microscope images of 3 different metasurface pixels, whereas for each structure, the meta-atoms have different lateral size to impose different phase delays to the light.

7.5 Characterization of SSE

7.5.1 Experimental setup

The experimental setup for characterizing the fabricated SSE, i.e. measuring the transmission spectrum of each pixel, is demonstrated in Figure 7.5. Broadband light (Thorlab High-Power Stabilized Quartz Tungsten-Halogen Source: $360 - 2500$ nm, SLS302) is coupled to an optical fiber, then collimated by a tube lens (Thorlab AC254-030-A-ML, $f = 30$ mm) at the fiber output and subsequently focused by an objective onto one of the pixels in the SSE. Fiber with core diameter: 10 μm (Thorlabs M64L01, 0.1NA), the objective: Nikon LU Plan Fluor 10x, $WD = 17.5$ mm, $NA = 0.30$, the beam size on the SSE: ≈ 20 μm. A 50/50 beam splitter is placed in between the collimator and the objective, combined with a tube lens (Thorlabs AC254-100-A-ML, f=100mm) and a camera (Thorlabs CS165CU) for viewing the patterns on the mask. The light transmitted through the SSE device is coupled to another fiber (Thorlabs M74L05, core diameter = 400 μm, 0.39 NA), which is connected to a spectrometer (Teledyne Princeton Instruments, IsoPlane SCT 320) for spectral measurements of light coupled to the fiber.

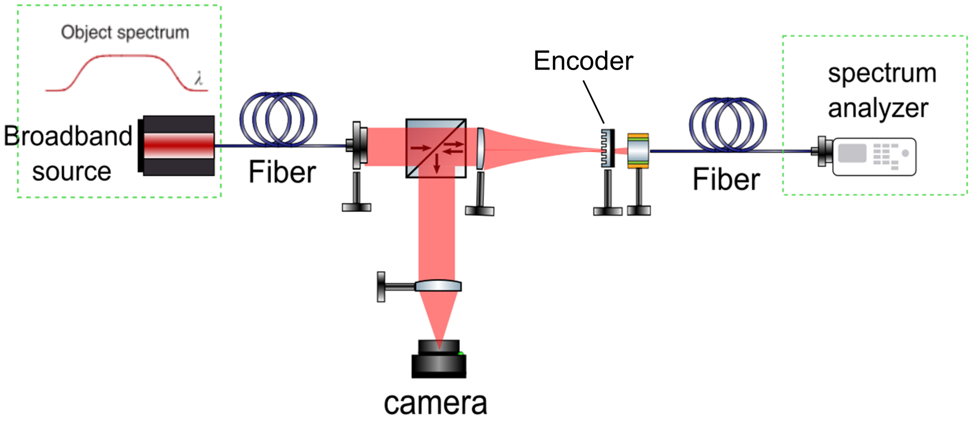

FIGURE 7.5: Experimental setup for measuring the transmission spectra of all the pixels in the encoder.

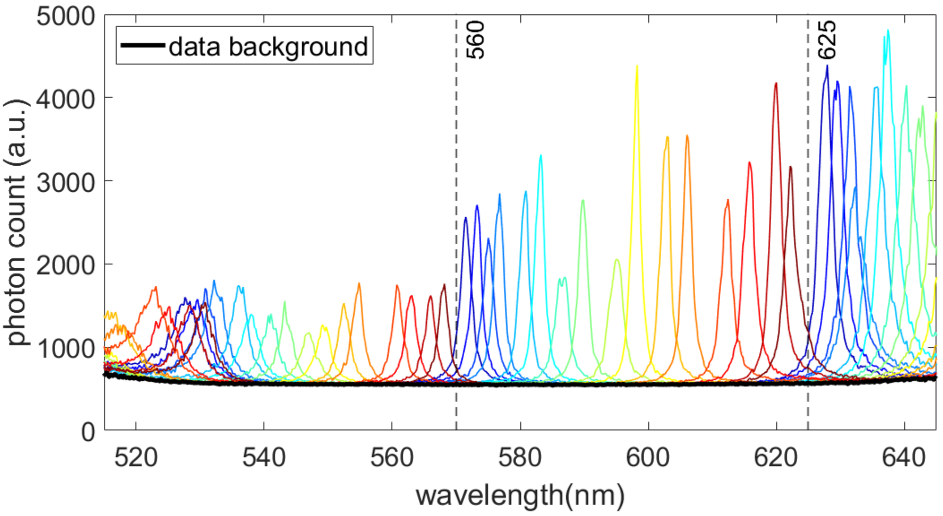

FIGURE 7.6: Measured transmission spectra of all the pixels of the encoder, displayed in different colors from blue to red.

7.5.2 Characterization results

As shown in Figure 7.6, the transmission resonance bands of each pixel in the SSE are spectrally shifted and distributed across the entire free spectral range of the cavity, owing to the phase shifts imparted by the metasurfaces. These transmission bands appear as narrow peaks with minimal spectral overlap, resulting in a low condition number of the encoding matrix—even without any device-level optimization.

For spatial-to-spectral encoding, we select transmission resonances within the spectral range of 570 nm to 625 nm, corresponding to the 9th resonance mode of the Fabry-Pérot cavity with a length of 2.5 µm. The full width at half maximum (FWHM) of the resonance peaks ranges from 1.6 nm to 3.0 nm, with an average value of approximately 2.3 nm, indicating a quality factor (Q) of about 30. Resonance modes of other orders are not utilized, as some lie near the edge of the DBR's stopband, where the cavity exhibits significantly lower Q factors and reduced spectral orthogonality among the encoding channels.

7.6 Spectrally encoding imaging experiments

7.6.1 Experimental setup

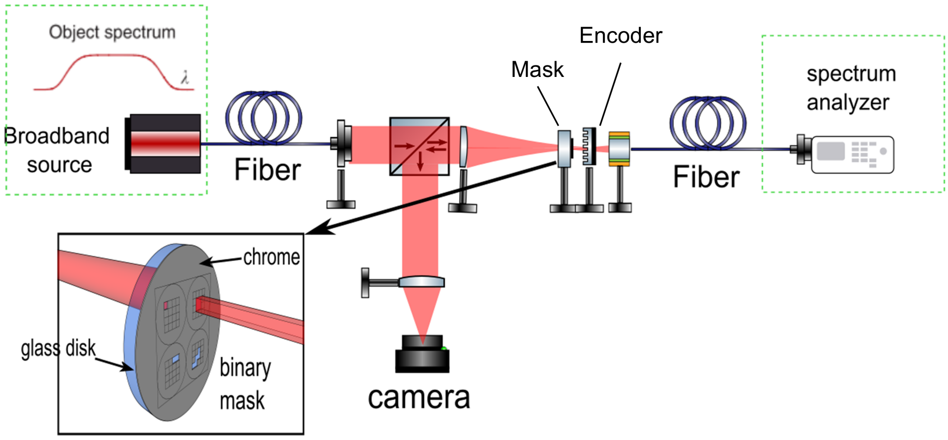

FIGURE 7.7: Experimental setup for the spectrally encoding imaging.

The fabricated SSE is used to encode light passing through various binary patterns into distinct spectra. The spectrally encoded light is then transmitted through an optical fiber and measured using a spectrometer. The recorded spectra are subsequently decoded to reconstruct the original spatial patterns.

The experimental setup is illustrated in Figure 7.7. This setup is similar to the characterization configuration, with the key difference being the placement of a metal mask with binary patterns in close proximity to the SSE, thereby generating spatial patterns to be spectrally encoded. A multimode fiber (Thorlabs M122L02, core diameter: 200 µm, NA = 0.22) delivers light from the source. The imaging objective used is a Nikon LU Plan Fluor 5× (NA = 0.15, WD = 23.5 mm), which results in an expanded beam spot of approximately 800 µm on the mask, ensuring full coverage of the pattern region.

7.6.2 Calibration of the encoding matrix

The encoding matrix $\mathbf{M}$ must be calibrated prior to imaging. This matrix is constructed using the measured transmission spectra of all spatial pixels in the SSE over the wavelength range of 570 nm to 625 nm. Because light may leak from the edges of the metal mask, carrying the transmission spectrum of the Fabry-Pérot cavity without a metasurface, we also include this background spectrum in the encoding matrix. The transmission spectra of all pixels and the background are plotted in Figure 7.8a.

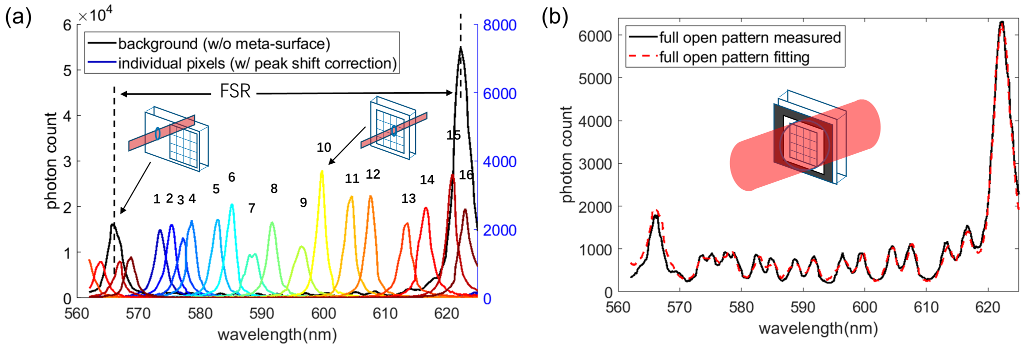

FIGURE 7.8: **(d)** The measured transmission spectra of all pixels of the encoder in different colors from blue to red. The transmission spectrum of the cavity without metasurface (the background) is also plotted (in black) for comparison. **(e)** The measured and fitting spectra of the all-open pattern. The fitting spectrum is the weight summation of the spectra of all pixels.

We calibrate the spectral code matrix **M** by fitting the measured spectrum of an all-open pattern, $I_0(\lambda)$, using the spectra of the individual pixels, $t_j(\lambda)$, and the background spectrum $t_{\text{bg}}(\lambda)$. The calibration procedure is defined as follows:

$$a_0 = \left[I_0(\lambda)\right]^T \tag{7.9a}$$

$$M = \left[c_{\text{bg}}, t_{\text{bg}}(\lambda+\Delta\lambda_0) \quad c_1, t_1(\lambda+\Delta\lambda_1) \quad \dots \quad c_{16}, t_{16}(\lambda+\Delta\lambda_{16})\right] \tag{7.9b}$$

$$b_0 = \begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}^T \tag{7.9c}$$

$$\min_{c_j,\Delta\lambda_j} |Mb_0 - a_0|^2 \tag{7.9d}$$

Here, $\Delta\lambda_j$ represents the wavelength shift applied to each spectrum, and c_j denotes the corresponding weighting factors.

The fitted spectrum of the all-open pattern, $I_{0\text{-fit}}(\lambda) = \mathbf{M}b_0$, is plotted in Figure 7.8b (red dashed line), which matches well with the measured spectrum $I_0(\lambda)$ (black solid line).

7.6.3 Imaging results

Using the calibrated encoding matrix **M** and the decoding method described in Chapter 7.2.2, we recovered the binary intensity patterns from the measured spectra of 24 different original patterns. The threshold for grey-scale to binary is set at the midpoint of the minimal and maximal greyscale values. Figure 7.9a and b shows two examples of the binary pattern decoding process with the SSE. The corresponding measured and fitted spectra of these two patterns are plotted in Figure 7.9c and d. The imaging process for all 24 patterns are summarized in Figure 7.10 and Figure 7.11.

To quantitatively evaluate the fidelity of the imaging approach, we define an error rate as the number of pixels that are incorrectly reconstructed divided by the total number of pixels. The error rates for all binary patterns are plotted in Figure 7.12a. We used a high resolution (0.2 nm) spectrometer, with $n = 316$ spectral data points in the wavelength range of $562 - 625$ nm (20 spectral data points per spatial pixel). This results in an average error rate of 9.8%. We further investigated the dependency

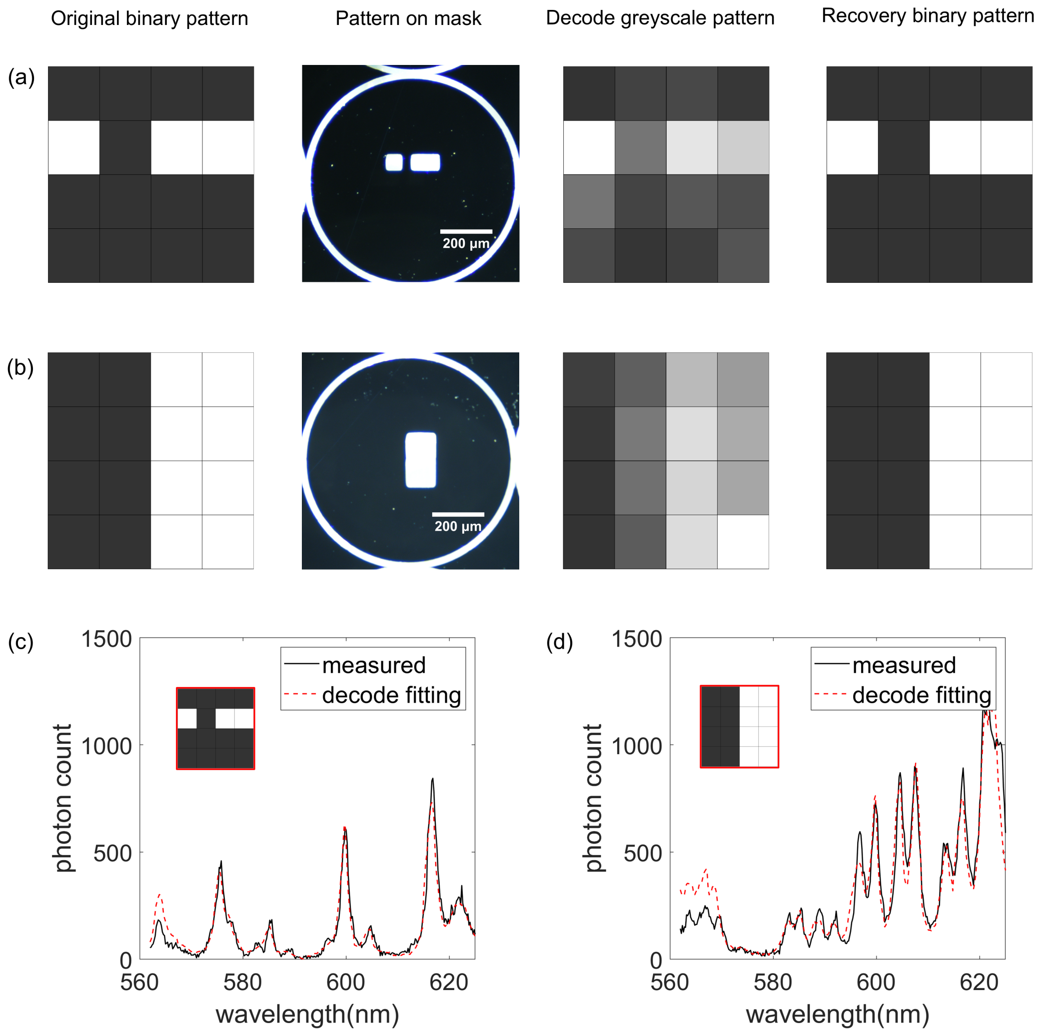

FIGURE 7.9: **(a)(b)** The 12th, 22nd original binary patterns, the microscope image of the corresponding binary patterns on the chrome mask, the decoded grey-scale patterns, and the recovered binary pattern from the decoded grey-scale patterns. **(c)(d)** The measured and fitted spectra of the 12th, 22nd patterns.

of the decoding error rate on the resolution of the pattern spectra. We synthesized low resolution spectral data by artificially compressing the number of data points in wavelength in the measured pattern spectra and encoding matrix, and repeated the computational decoding on these low resolution pattern spectra. As can be seen in Figure 7.12b, the decoding error rate only marginally increases when the spectral resolution decreases. As long as the resolution is better than 1.5 nm the error rate is maintained below $\sim$ 11%. This shows that with our SSE, the decoding can be efficiently performed using a low resolution spectrometer.

While most patterns are accurately reconstructed (one or less incorrect pixel assignment), some (particularly, 9th and 13th, as shown in Figure 7.10) yield extremely high error rates, which contributed most to the $\sim$ 10% average error rate of the spatial-spectral decoding experiment. This high error rate may stem from the divergence of light transmitted through the pattern. Our SSE device is based on the assumption that the incident light is a plane wave. This assumption is satisfied at near-field imaging condition, where the Fresnel number, $F = a^2/\lambda L \gg 1$. In our experiment

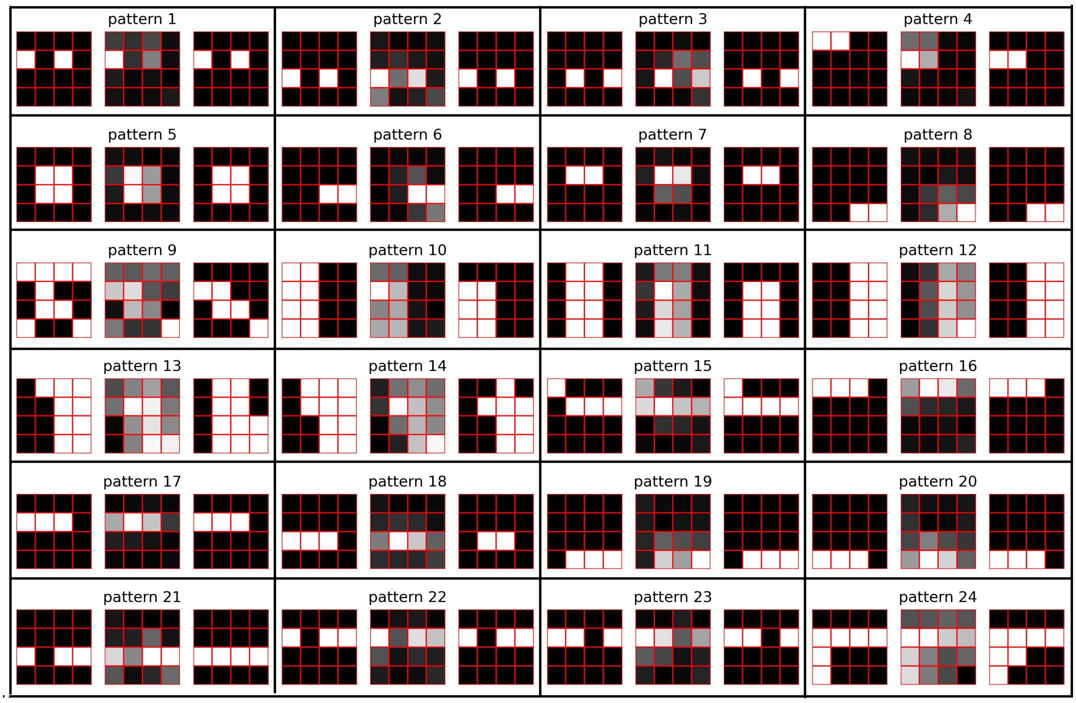

FIGURE 7.10: Original binary patterns, the decoded grey-scale patterns, and the recovered binary pattern for all 24 patterns.

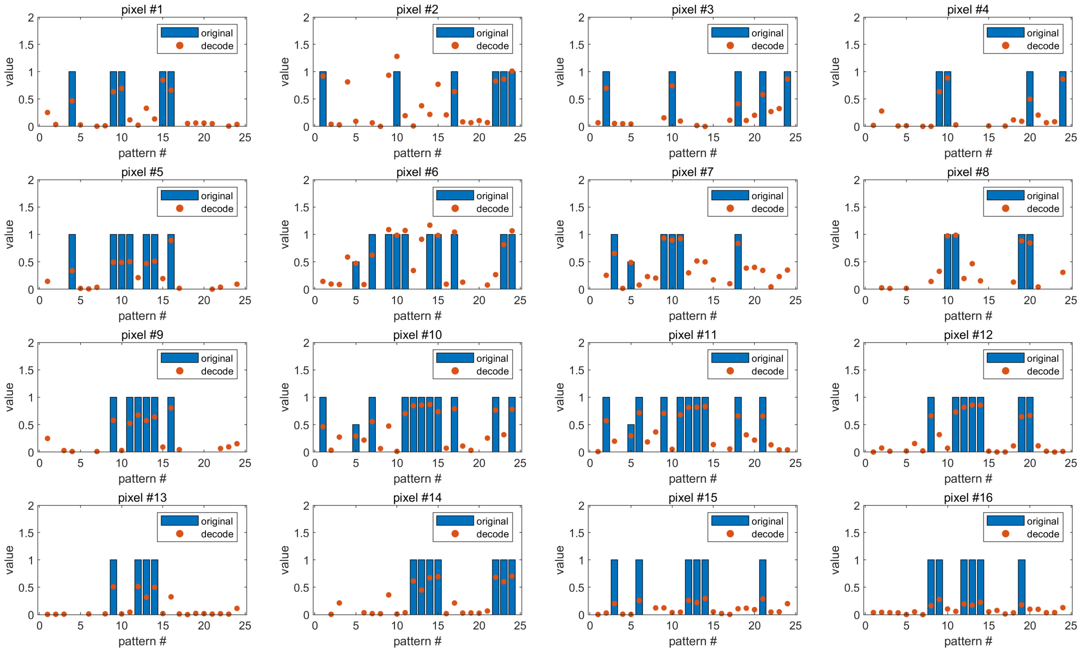

FIGURE 7.11: The original and decoded grey-scale values of all the pixels for all 24 patterns.

setup, the characteristic length for a single pixel of the pattern is half the size of the pixel, $a = 25\,\mu\text{m}$, the propagation distance $L \approx 600\,\mu\text{m}$ (slightly thicker than the substrate 500 μm), and the wavelength $\lambda \approx 560\,\text{nm}$. These parameters give a Fresnel number $F \approx 1.8$, which indicates that we expect the light beam to diverge. Essentially, the light from one single pixel of the pattern on the mask could leak to the surrounding pixels on the device.

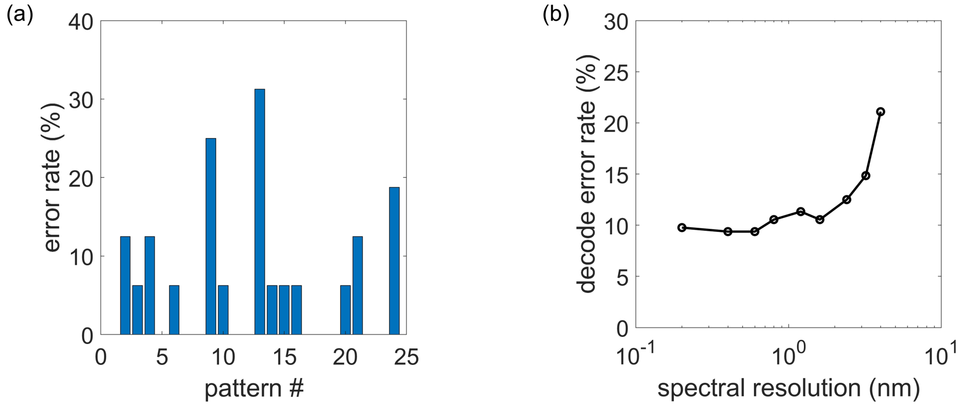

FIGURE 7.12: **(a)** The error rates of the 24 recovered binary patterns compared to the original patterns, using the high resolution (0.2 nm) measured spectra data, which contains ≈ 20 spectrum datapoints per pixels. **(b)** The pattern recovery average error rates (over 24 patterns) as a function of the resolution of the pattern spectra.

7.7 Discussion

We have demonstrated cavity–metasurface-based spectrally encoded imaging. The fabrication of the encoder is fully compatible with standard semiconductor processes, making it well-suited for large-scale production. However, each encoder still requires an individual calibration procedure to account for fabrication-induced variations and alignment with other optical components. Once calibration is completed, the subsequent computational decoding reduces to a simple matrix multiplication, enabling highly efficient implementation.

Although the small pixel number (16) of our SSE device limits its application for conventional imaging, some laser guided surgeries require as little as 3 pixels to navigate the movement of surgical instruments [31]. Thus, an SSE device with only 16 pixels can already be useful and potentially replace the previously used multi-core fibers for collecting such information, while being thinner, cheaper, more robust, and easier to integrate.

The number of spatial pixels in the current SSE is fundamentally limited by the FSR of the Fabry-Pérot cavity relative to the resonance bandwidth of each filter—that is, by the finesse of the cavity. To overcome this limitation, the metasurface itself can be engineered to perform spectral filtering, thereby enabling the construction of an SSE without relying on a Fabry-Pérot structure, as illustrated in Figure 7.13. When meta-atoms are placed in close proximity, the presence of resonance modes arising from their collective interaction lead to enrich spectral responses [32], as shown in Figure 7.13. This approach allows for a diverse set of spectral responses that are not constrained by the cavity finesse. Moreover, a purely metasurface-based design simplifies the fabrication process and facilitates easier integration with miniaturized imaging systems such as endoscopes.

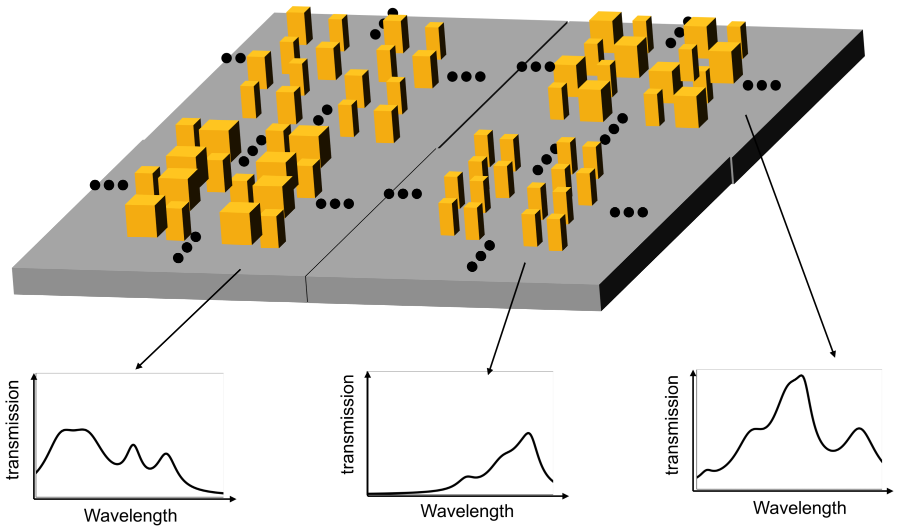

FIGURE 7.13: Schematic demonstration of a pure metasurface spatial-to-spectral encoder.

Chapter 8

Hyperspectral imaging enabled by metasurface code mask

8.1 Overview

An array of spectral filters, in addition to serving as a spatial-to-spectral encoder for spectrally encoded imaging as discussed in Chapter 7, can also function as a code mask for compressive sensing–based hyperspectral imaging. Hyperspectral imaging captures spectral information at every spatial pixel of a scene and has wide-ranging applications in precision agriculture, environmental monitoring, and medical diagnostics. However, conventional camera sensors can only record light intensity, lacking inherent spectral resolution. By placing a code mask in front of the sensor, hyperspectral information can be encoded into spatial intensity variations, enabling indirect spectral capture. The hyperspectral information can be later reconstructed computationally. Unlike spectrally encoded imaging, the reconstruction is an underdetermined problem. However, assuming that the spatial and spectral signals are sparse in the hyperspectral data, compressive sensing can be applied to reconstruct the full hyperspectral cube.

In this chapter, we present a metasurface-based code mask designed for compressive sensing based hyperspectral imaging. The mask is optimized using an iterative algorithm inspired by thermodynamic processes to minimize reconstruction error. It enables the encoding of a full hyperspectral dataset—comprising 140×200 spatial pixels and 21 spectral bands spanning 480–680 nm—into a single grayscale image of size 140×200, achieving a 21-fold data compression. From this single measurement, the hyperspectral information is computationally reconstructed.

We validate the system by capturing high-resolution, snapshot hyperspectral images of natural objects under broadband illumination, demonstrating both its practicality and potential for real-world deployment. Compared to most commercial hyperspectral imagers based on tunable filters, line-scan spectrometry, or multi-aperture architectures, our metasurface-based imager offers simplified configuration and superior performance in terms of imaging speed, resolution, and optical efficiency.

We credit our collaborator, Prof. Vishwanath Saragadam and his group members at the University of California, Riverside for performing computational reconstruction of the hyperspectral images based on compressed sensing algorithms.

This chapter is adapted from an unpublished paper titled "Snap-shot hyperspectral imaging enabled by metasurface".

8.2 Conventional hyperspectral imager

A hyperspectral image with $H \times W$ spatial pixels and K spectral bands can be denoted as $I(x, y, \lambda_i)$, where x and y are the indices of the spatial pixels, and i is the index of the spectral bands, with $x, y, i \in \mathbb{Z} \mid 1 \leq x \leq M, 1 \leq y \leq N, 1 \leq i \leq K$. Since a conventional camera sensor can only record the intensity of incident light, it cannot directly capture the full $H \times W \times K$ hyperspectral data cube in a single snapshot. Traditional hyperspectral imaging systems either employ a scanning mechanism to sequentially acquire different portions of the cube or utilize mosaic-based architectures, which expand each spatial pixel into K sub-pixels, requiring a camera with $M\sqrt{K} \times N\sqrt{K}$ pixels to fully record the hyperspectral data.

In scanning-based hyperspectral imaging systems, one common approach is to use a tunable filter to sequentially scan across wavelengths λ_i, acquiring intensity images $I_{\lambda_i}(x, y)$ at each spectral band [33, 34, 35, 36]. Alternatively, line-scan spectrometry combines a scanning slit with a transverse dispersive element, allowing the sensor to capture $I_y(x, \lambda_i)$ while scanning along the y direction [37, 38, 39]. Although these systems offer high spatial and spectral resolution, their imaging speed is fundamentally constrained by the sequential nature of the acquisition. Additionally, the integration of tunable filters, scanning slits, and dispersive elements leads to increased system complexity and size.

In contrast, mosaic-based systems incorporate narrowband multispectral filter arrays directly onto the $H \times W$ pixel sensor, enabling simultaneous acquisition of K spectral bands through optical filtering [40, 41]. These parallel acquisition systems significantly enhance imaging speed and enable the development of compact, snapshot-capable hyperspectral cameras. However, this comes at the cost of reduced spatial resolution, which is directly traded off to increase spectral resolution. Furthermore, systems relying on narrowband filters inherently face a trade-off between the number of spectral bands and photon collection efficiency. Specifically, an N-channel hyperspectral imager captures only $1/N$ of the available photons, which significantly limits performance in low-light or dark environments.

8.3 Compressive sensing hyperspectral imager

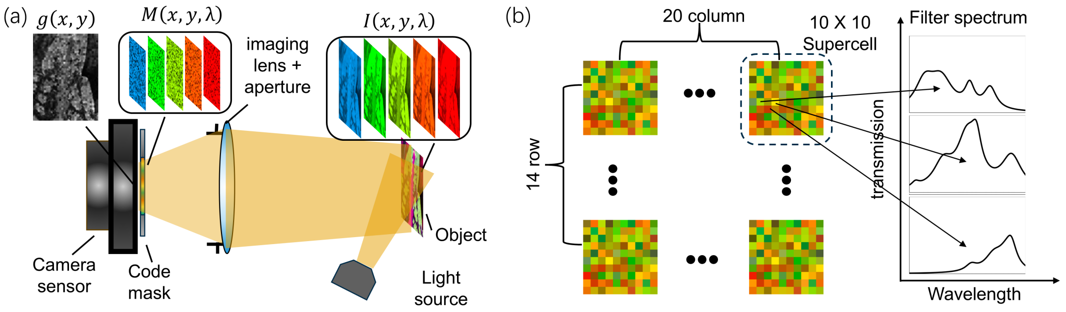

FIGURE 8.1: **(a)** Schematic of the hyperspectral imaging system. **(b)** Schematic of the spectral code mask for hyperspectral imaging, which contains an array of 20×14 repetitive supercells. Each supercell contains an array of 10×10 different spectral filters with distinct transmission spectrum. The color and the transmission spectrum of the filters are randomly generated only for demonstration.

Recently, compact hyperspectral imaging has been demonstrated with spectral code masks. These, when combined with recent advances in compressive sensing [42, 43], have offered a way to achieve snapshot hyperspectral imaging without dramatically sacrificing spatial resolution [44, 45, 46, 47].

As demonstrated in Figure 8.1a, the code mask is placed on top of the camera sensor with pixel-to-pixel alignment. Such a code mask, whose optical functionality is described by $\mathbf{M}(x, y, \lambda_i)$, is also a spectral filter array, with the filters at different location (x, y) having unique transmission spetra $t_{\mathrm{xy}} = \mathbf{M}(x, y, \lambda)$. However, unlike the mosaic-based system where $\sqrt{K} \times \sqrt{K}$ sub-pixels are used to record K bands of spectral information, the code masks compress the $H \times W$ pixels $\times K$ spectral bands hyperspectral data cube $I(x, y, \lambda)$ into a $H \times W$ pixels grayscale image, $g(x, y)$:

$$g(x,y) = \sum_{i=1}^{K} \mathbf{M}(x, y, \lambda_i) I(x, y, \lambda_i) \tag{8.1}$$

The grayscale image $g(x, y)$ captured by the camera sensor contains only $H \times W$ data points, representing an under-sampled observation of the full hyperspectral data cube $I(x, y, \lambda)$ with dimensions $H \times W \times K$. Reconstructing $I(x, y, \lambda)$ from $g(x, y)$ is thus an ill-posed inverse problem. However, this inversion becomes tractable by leveraging the inherent signal sparsities in hyperspectral data, as the true information content in $I(x, y, \lambda)$ is significantly lower than its raw dimensionality suggests.

First, natural scenes typically consist of a few distinct objects with well-defined edges rather than random textures across the field. As a result, the spatial content of the image is compressible and can be represented with far fewer coefficients in an appropriate spatial transform, much less than the full $H \times W$ pixel count, known as spatial sparsity.

Second, the spectral signatures of real-world materials tend to be smooth and structured rather than exhibiting random or spiky variations. This means that each pixel's spectrum is sparse or compressible in a suitable spectral basis, reducing the need for K independent values to describe each spectrum, which is known as spectral sparsity.

Third, spatially neighboring pixels often correspond to the same object or material and thus exhibit similar spectral profiles. This introduces redundancy across both the spatial and spectral dimensions of the data cube, rendering it highly compressible. Such internal redundancy is commonly referred to as global low-rankness [48].

To exploit these signal sparsities, we first reshape the spatial domain into a vector of dimension HW, reformulating the hyperspectral cube $I(x, y, \lambda)$ as a matrix $\mathcal{H} \in \mathbb{R}^{HW \times K}$. We then model this matrix as a low-rank product $\mathcal{H} = UV^{\top}$ [49], where $U \in \mathbb{R}^{HW \times r}$ captures the spatial components (left singular vectors), and $V \in \mathbb{R}^{K \times r}$ encodes the spectral signatures (right singular vectors). To promote spatial sparsity and smoothness, we replace U with an implicit neural representation (INR) using wavelet-based activations, which acts as an untrained neural prior that captures edges and textures while suppressing noise [50]. Finally, we impose a total variation (TV) regularization on U to encourage spatial smoothness and a ridge penalty on V to promote smooth spectral variations.

We present a hyperspectral imaging system based on spectral code mask and compressive-sensing. As a proof-of-concept, we design a code mask consisting of 140 × 200 spectral filters that work in 21 spectral bands spanning a range of 480 nm - 680 nm with $\lambda_i = 480 + 10i$ for $i = 0, ..., 20$. This spectral range is only limited by the tunable range of our tunable laser, which is used for the calibrations and benchmark testings of our imager. As shown in Figure 8.1b, the 140 × 200 array contains 20 columns and 14 rows of repetitive supercells. The supercell contains a 10 × 10 array of 100 different spectral filters, where each filter corresponds to one pixel and has a unique broadband transmission spectrum.

8.4 Metasurface spectral filters

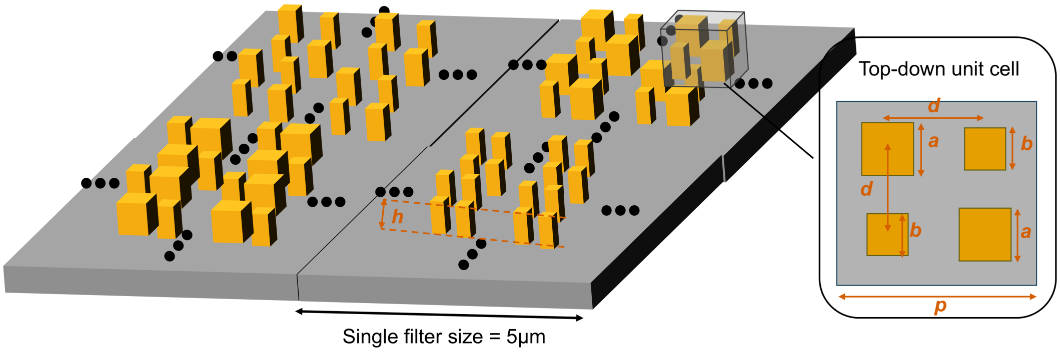

FIGURE 8.2: The schematic of a metasurface spectral filter that contains periodic pillar structures. The materials of the pillar and the substrate are Si and Al_2O_3, respectively. The height of the pillar is $h = 220$ nm. The 4-dimensional parameters p, d, a, b are tunable parameters for producing different spectra.

The spectral filters in the code mask for hyperspectral imaging can be implemented either using Fabry-Pérot cavities with varying thicknesses [45] or materials with diverse absorption spectra [46]. However, the fabrication of code masks in these approaches typically requires multi-stage lithography processes, which introduce significant complexity in fabrication and hinder low-cost, large-scale industrial production. Moreover, the spectral filters based on optical cavities or absorption materials often exhibit limited tunability, thereby constraining the flexibility available for optimizing the code mask design.

A metasurface that consists of a periodic array of subwavelength meta-atoms is a promising candidate for these spectral filters [51, 52]. The resonance modes induced by neighboring meta-atoms result in transmission spectra that are sensitive to the geometries of the meta-atoms. Therefore, tuning the meta-atom geometries enables us to implement filters with diverse spectral responses that have a high degree of linear independence, which is essential to achieve high resolution spectra at each spatial pixel. In addition, a metasurface can be fabricated with a simple one-step lithography process, enabling the large-scale and cost-effective fabrication of a dense array of filters with a pixel size of only a few micrometers, matching the typical pixel density in a camera sensor.

Therefore, we design a code mask using metasurface spectral filters. These metasurfaces have a periodic crystalline silicon-on-sapphire pillar structure. As can be seen in Figure 8.2, the unit cell of the periodic structure contains four pillars in a 2x2 grid, with diagonal pillars being identical. The sizes of the pillars a, b, the distance between the pillars d, and the periodicity of the unit cells p are the four geometric parameters which can be tuned to realize different transmission spectra. This configuration of the unit cell is chosen because it is polarization independent and has a large degree of tunability, while the shape of the pillars has a low level of complexity, reducing the fabrication imperfections and lithography write time. The size of each single filter is 5 µm.

8.5 Design of code mask

The 100 distinct filters forming the supercell for the code mask are optimized to ensure a low spatial and spectral correlation between the code mask and the hyperspectral data, thereby reducing the reconstruction error in the compressive sensing [45]. To guide this optimization, two figures of merit are defined for reflecting how good the code mask is for reducing the reconstruction error: the spatial randomness R_i for each wavelength band i, and the spectral correlation coefficients C_{ij} between every pair of spectral bands i and j. To calculate these two figures-of-merit, the transmission of the code mask in each wavelength band, $\mathbf{M}(x, y, \lambda_i)$, is flattened into a one-dimensional vector $\mathcal{T}_i$ that contains the transmittance values for all pairs of (x, y):

$$\mathcal{T}_i = [t_{i1}, \cdots, t_{ik}, \cdots, t_{i100}] \tag{8.2a}$$

$$t_{ik} = \mathbf{M}(x, y, \lambda_i) \tag{8.2b}$$

$$\text{where } k = 10(y-1) + x, 1 \leq k \leq 100, 1 \leq x, y \leq 10; k, x, y \in \mathbb{Z} \tag{8.2c}$$

The spatial randomness R_i is defined as σ_i/μ_i, where σ_i and μ_i are the standard deviation and mean values of all elements in $\mathcal{T}_i$. The spectral correlation coefficient C_{ij} is the absolute value of the Pearson correlation coefficient between the two vectors $\mathcal{T}_i$ and $\mathcal{T}_j$. Specifically:

$$\mu_i = \frac{1}{100} \sum_{k=1}^{100} t_{ik} \tag{8.3a}$$

$$\sigma_i = \sqrt{\frac{1}{100} \sum_{k=1}^{100} (t_{ik} - \mu_i)^2} \tag{8.3b}$$

$$C_{ij} = \left| \frac{\sum_{k=1}^{100} (t_{ik} - \mu_i)(t_{jk} - \mu_j)}{100 \sigma_i \sigma_j} \right| \tag{8.3c}$$

$$R_i = \sigma_i / \mu_i \tag{8.3d}$$

Large values for R_i and small values for off-diagonal C_{ij} (with $i \neq j$) are desired for reducing the reconstruction error.

8.5.1 Code mask optimization framework

The code mask is optimized by selecting 100 different filters from a library containing $\sim$ 2000 filters to construct the supercell, aiming at simultaneously reducing C_{ij} $(i \neq j)$ and increasing R_i of the code mask. The filter library is built by scanning the geometric parameters of the meta-atoms within a certain range governed by fabrication feasibility, from which 100 filters are selected. The spectra of these filters in the library are experimentally characterized (see the following chapter 8.5.3)

This selection is performed through an iterative optimization process that minimizes the loss function F defined from R_i and C_{ij} $(i \neq j)$ (see the following chapter), as illustrated in Figure 8.3b. Initially, 100 filters are randomly chosen from the library to form a starting configuration. The loss function is then computed based on the experimentally measured spectra of the selected filters. At each iteration, one filter within the supercell is randomly selected and proposed for replacement with another filter randomly drawn from the remaining unselected filters in the library. The new configuration's loss function is recalculated and compared with the previous value. If replacement leads to a lower loss function, it is accepted, and the supercell is updated accordingly.

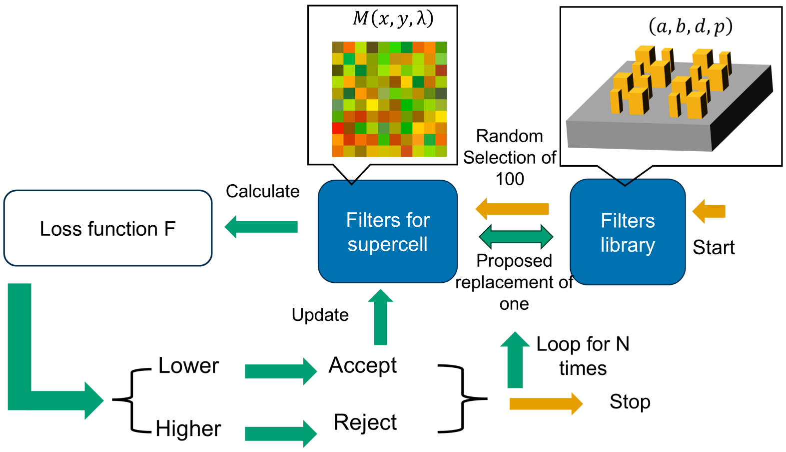

FIGURE 8.3: A schematic diagram demonstrating the iterative optimization of the code mask.

8.5.2 Loss function

The loss function F is computed from the spectral correlations C_{ij} between every two spectral bands i and j, and the spatial randomness R_i for each spectral band i:

$$F = H_4(\{C_{ij}\}) - \alpha \bar{R} \tag{8.4}$$

where $H_4(\{C_{ij}\})$ is the term that counts contributions from spectral correlation coefficients C_{ij}, and $\bar{R}$ counts contributions from spatial randomness R_i. α is a tunable coefficient that balances the trade-off between spectral correlation and spatial randomness.

$$H_4(\{C_{ij}\}) = \frac{1}{21}\sum_{i=0}^{20} \|\mathbf{c_i}\|_4 \tag{8.5}$$

$$\mathbf{c}_i = \left[C_{ij}\right]_{j=0,j\neq i}^{20} \tag{8.6}$$

$$\bar{R} = \frac{1}{21}\sum_{i=0}^{20} R_i \tag{8.7}$$

8.5.3 Metasurface filter library

As shown in Figure 8.2, the unit cells of the metasurface filters contain two pairs of identical square pillars located at the diagonal corners, with four geometric parameters p, d, a, b. a and b are the size of the pillars, d is the spacing between two different pillars, and p is the size of the unit cell. These four geometric parameters are scanned across the following range with a step of 16 nm to form a library containing around 2000 filters with different transmission spectra:

$$304\,\mathrm{nm} \leq p \leq 500\,\mathrm{nm} \tag{8.8}$$

$$96\,\mathrm{nm} \leq d \leq 0.5p \tag{8.9}$$

$$64\,\mathrm{nm} \leq b \leq a \leq d \tag{8.10}$$

$$a + b \leq 2d - 64\,\mathrm{nm} \tag{8.11}$$

We fabricate a chip containing all of these 2000 filters using electron beam lithography. The process is the same as shown in Chapter 5.3. The transmission spectra of these 2000 filters are then experimentally measured using the setup demonstrated in Figure 8.4.

As indicated, the image of the metasurface spectral filter array chip is relayed by a 10× relay optics (tube lens: TL200 from Thorlabs, focal length 200 mm, anti-reflection coating 350-700 nm; objective lens: Mitutoyo M Plan Apo 10x LWD Objective, with NA = 0.28, focal length = 20 mm, working distance = 34 mm) to a monochrome camera sensor (ZWO ASI183MM Monochrome Astronomy Camera). The metasurface filter array chip is placed at the focal plane of an imaging lens with tunable aperture (Edmund optics 50 mm C Series Fixed Focal Length Lens, the f-number is set as 2.8). A supercontinuous laser (NKT Photonics, SuperK Fianium FIR20, paired with SuperK SELECT 4× VIS/IR tunable filter) is used for illumination. The wavelength of the output light is scanned across $480\,\mathrm{nm} - 680\,\mathrm{nm}$. The image of the filter array is captured at each wavelength, and the spectra of all the filters can be extracted from these images. Figure 8.4 shows the images of the 2000 filters at 500 nm, 570 nm, and 640 nm.

Note that a frosted tape is placed ahead of the light source as a diffuser, and is blown by a fan. The dynamically moving tape vibrates such that the speckle effect can be eliminated by time averaging in the image capturing processes with exposure times as long as a few seconds. The noise level of the images captured by the camera sensor is evaluated by capturing the dark images (light source being turned off) with the same exposure time, and comparing the intensity of the dark images with the bright images. This noise level is found to be below 1% of the maximum intensity.

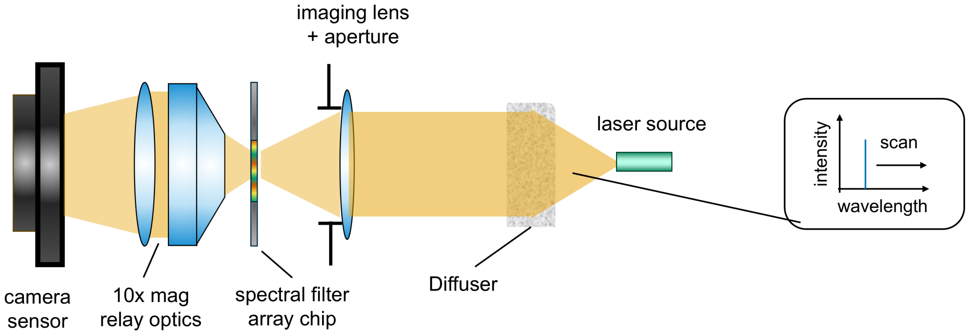

FIGURE 8.4: The setup for massively parallel measuring the transmission spectra of metasurface spectral filters.

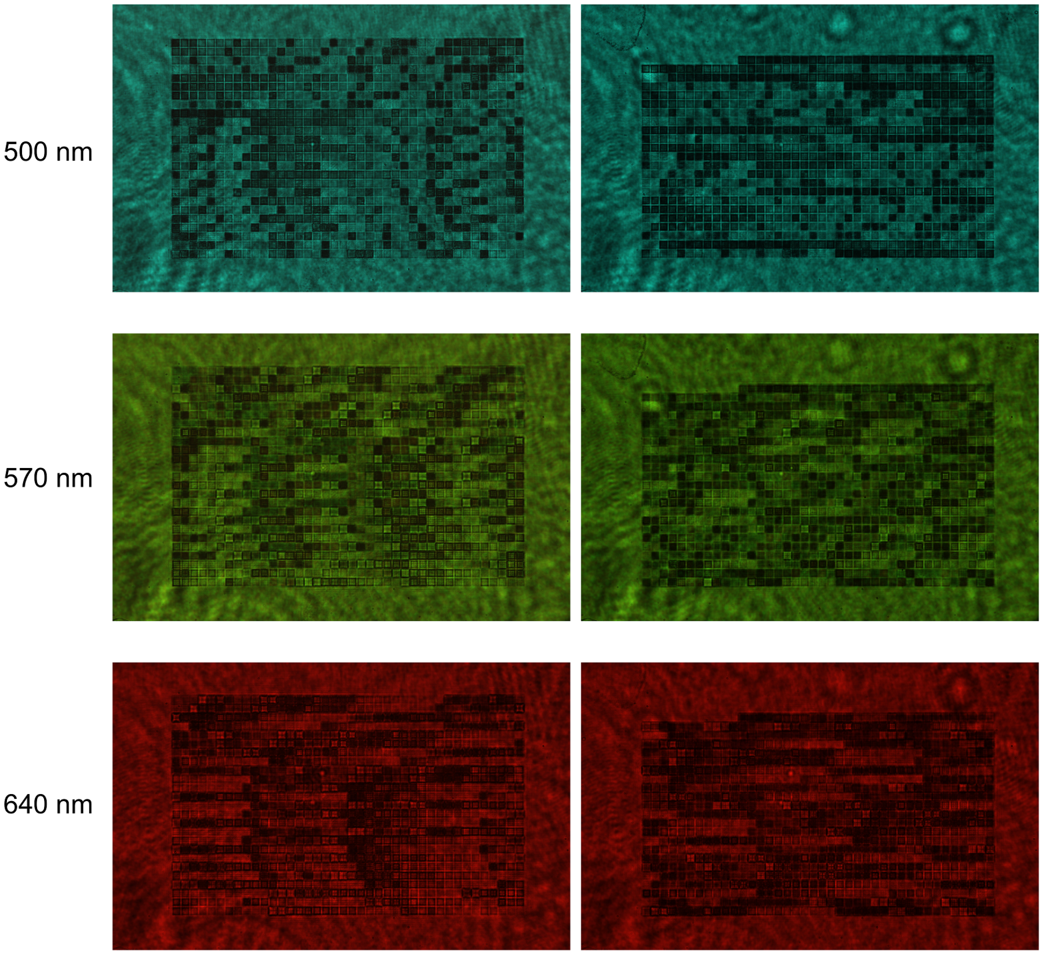

FIGURE 8.5: Transmissive images of the metasurface filter library at three different wavelengths.

8.5.4 Code mask optimization results

The iterative optimization described in chapter 8.5.1 loops for 20,000 iterations, which is sufficient to reduce the loss function to a plateau, beyond which further iterations yield no significant improvement, as shown in Figure 8.6b. Different α values result in different trade-offs between the spectral correlation indicator $H_4(\{C_{ij}\})$ and the spatial randomness indicator $\bar{R}$, as can be seen in Figure 8.6a. We set $\alpha = 0.7$ to reach a balance between reducing the spectral correlations and increasing the spatial randomness. Figure 8.6c shows the spectral correlation matrix of the supercell at the beginning (random selection) and after optimization. As can be seen, the off-diagonal spectral correlations is greatly reduced after optimization (average off-diagonal correlation $\overline{C_{ij}}|_{i \neq j}$ dropping from 0.53 to 0.24). Figure 8.6d shows the spatial randomness of the supercell in each spectral band at the start and after optimization, demonstrating that the optimization maintains the spatial randomness (average value is 0.36 at the start and 0.38 after optimization). Thus, this optimization of the supercell can reduce reconstruction errors in HSI imaging.

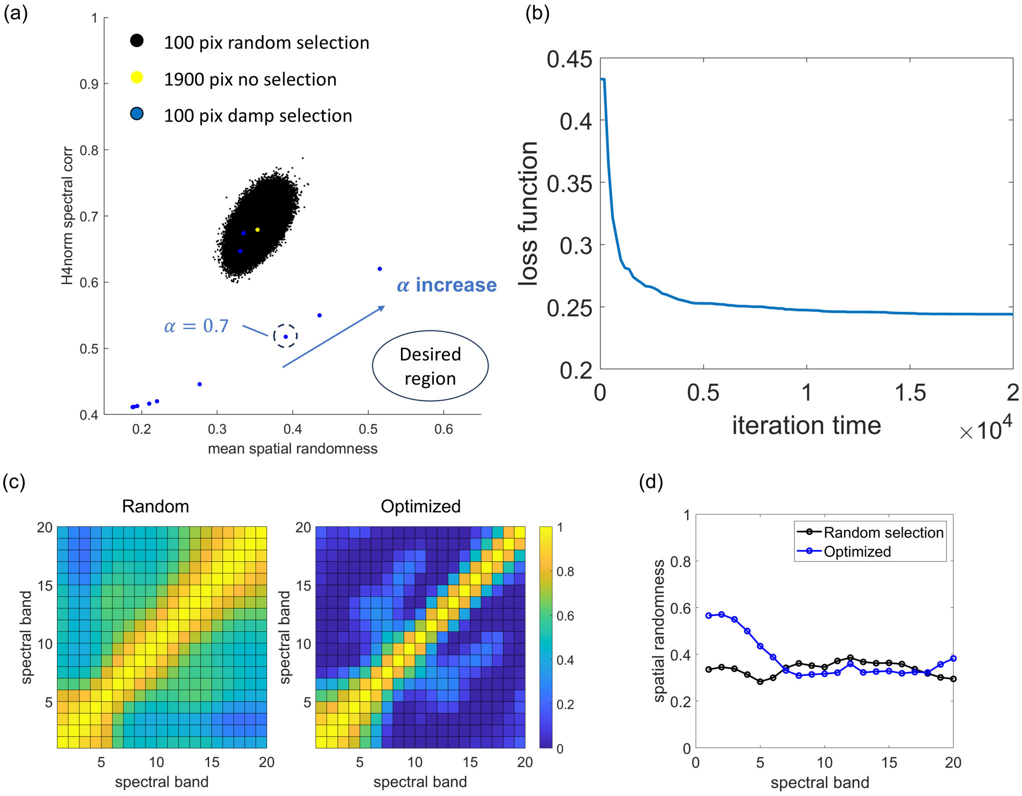

FIGURE 8.6: **(a)** FOMs of code mask configurations randomly generated and after optimization. **(b)** Loss function vs. iteration time during the optimization process, with $\alpha = 0.7$. **(c)** Spectral correlation matrix of the supercell in 21 different spectral bands across the 480 nm – 680 nm range with a step of 10 nm, before and after the optimization. **(d)** Spatial randomness of the code mask before and after the optimization.

8.5.5 Codemask calibration

Figure 8.7a shows the optical microscope image of the fabricated code mask under ambient light. As can be seen, different filters appear as blocks with different colors under ambient light illumination, indicating different optical responses. Figure 8.7b shows the SEM images of four of the metasurface filters in Figure 8.7a.

The fabricated code mask is calibrated using the same setup in Figure 8.4. The average transmission maps of the 10×10 filters supercell in 21 different spectral bands are shown in Figure 8.8a. The spectral correlation matrix and the spatial randomness of the code mask are plotted in Figure 8.8b and c, respectively. The off-diagonal

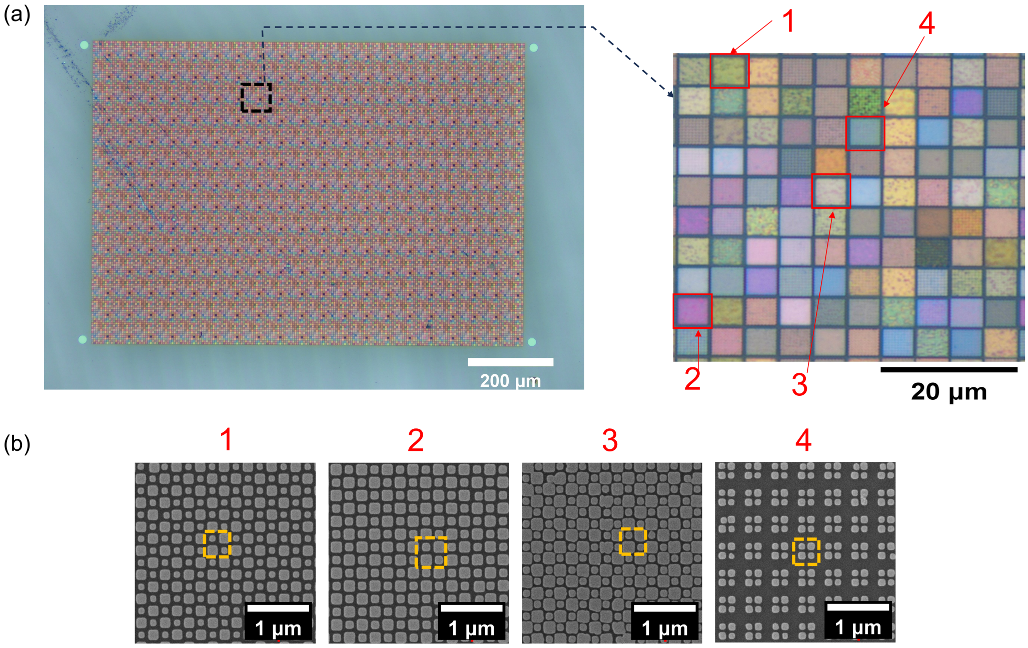

FIGURE 8.7: **(a)** Optical microscope image of the fabricated metasurface spectral code mask. The insertion is a zoomed-in view showing a single supercell containing 10 × 10 different filters, each of which has a size of 5 µm, and has a distinct transmission spectrum within the 480 nm – 680 nm wavelength range. **(b)** Top-down SEM images of the 4 metasurface pixels that are enclosed by the red lines in (a). The orange dashed line in the SEM images indicates a unit cell of the metasurface.

values of the spectral correlation coefficients C_{ij} ($i \neq j$) has an average value of 0.27, and the spatial randomness $R_i \geq 0.5$ for all spectral bands, with an average value of 0.56. These ensure a low reconstruction error. Figure 8.8d shows the measured photon efficiencies (average transmission μ_i) of the code mask across all spectral bands. The photon efficiencies range from 25% to 50%, with an average of 39%, greatly exceeding the theoretical limit of 5% for a narrow band filter based hyperspectral imager.

8.6 Hyperspectral imaging experiment

We test our hyperspectral imaging system by performing imaging experiments on transmission targets, color checker, and natural plant objects. These experiments demonstrate the spatial resolution and spectral accuracy of the imaging system.

8.6.1 Imaging transmission target

To evaluate the spatial resolution of the system, we first imaged a USAF resolution target illuminated by a tunable laser, as illustrated in Figure 8.9a. The hyperspectral imaging system consisted of a monochrome camera sensor, a 10× relay optics, the metasurface code mask chip, and an imaging lens with tunable aperture, as decribed in Chapter 8.5.3. In addition, a 480 nm–680 nm band-pass filter (combination of two filters: Semrock BrightLine Multiphoton Filter FF01-680/SP-25 and Semrock BrightLine Multiphoton Filter FF01-496/LP-25) is added to filter out the light outside of the working wavelength range. Light containing two distinct wavelengths passes through a diffuser and then transmits through the USAF target mask. Figure

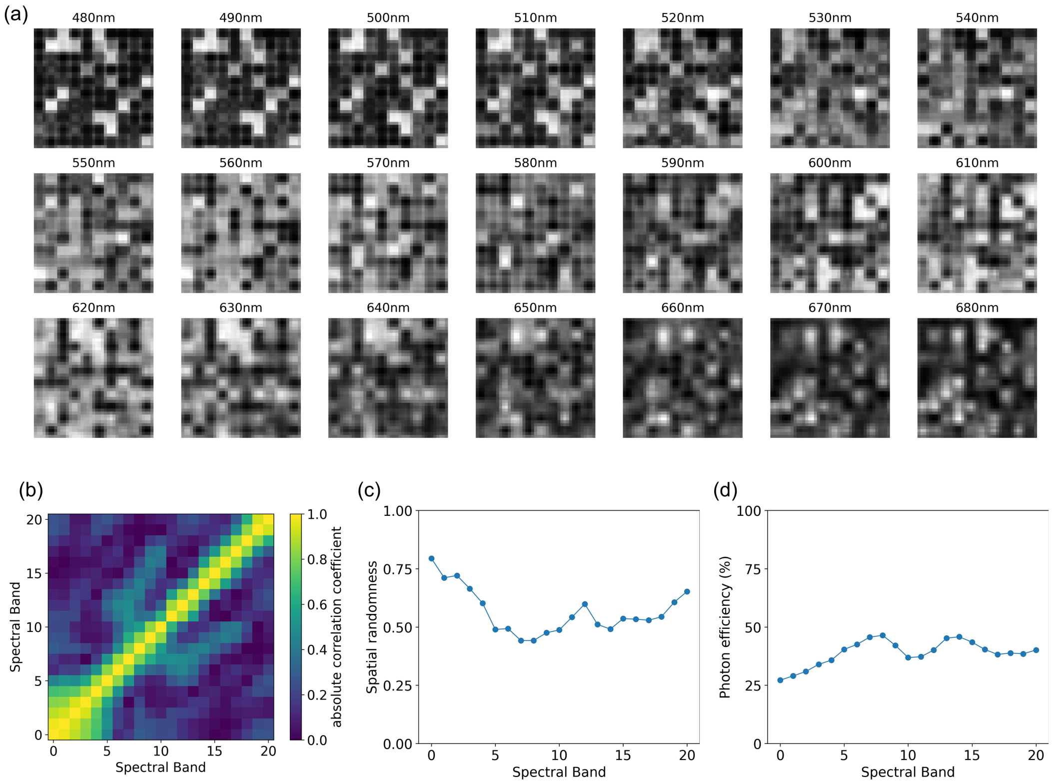

FIGURE 8.8: **(a)** Average transmission map of the 10 × 10 filters supercell in 21 different spectral bands across the 480 nm – 680 nm range. **(b)** Spectral correlation matrix of the fabricated code mask. **(c)** Spatial randomness of the fabricated code mask in 21 different spectral bands. **(d)** Photon efficiencies of the fabricated code mask in 21 different spectral bands.

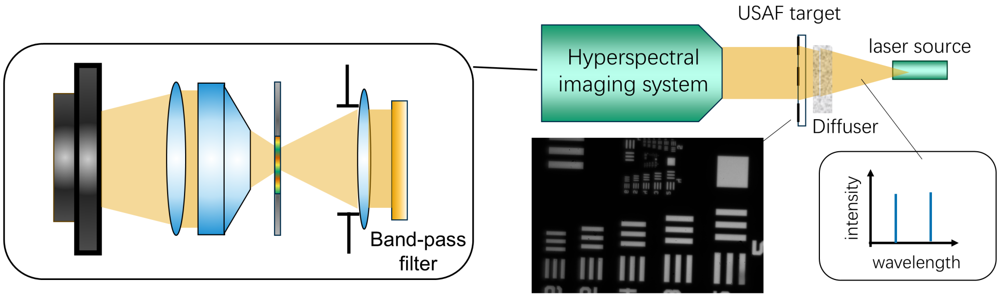

FIGURE 8.9: Experiment setup for imaging the USAF transmission target under laser illumination with two peak wavelengths. Lower middle insertion: grey scale image of the USAF target taken without the spectral code mask.

8.10a shows the compressed image of the target captured by our hyperspectral imaging system. The illumination consists of two wavelength components (510 nm and 570 nm). The reconstructed hyperspectral images exhibit two sharp spectral peaks corresponding to the illumination wavelengths, while in other bands, the average image intensity remains below 5% of the peak values, as shown in Figure 8.10b. This demonstrates a spectral resolution of 10 nm. The USAF target features are faithfully recovered at two illumination wavelengths, as shown in Figure 8.10c and Figure 8.10d. Figure 8.10e plots the image intensity profile across the grating pattern along the dashed white square in Figure 8.10d. As shown, two grating lines with a pitch of only two metasurface pixels are clearly resolved, confirming that the hyperspectral image reconstruction preserves the same spatial resolution as the imaging system.

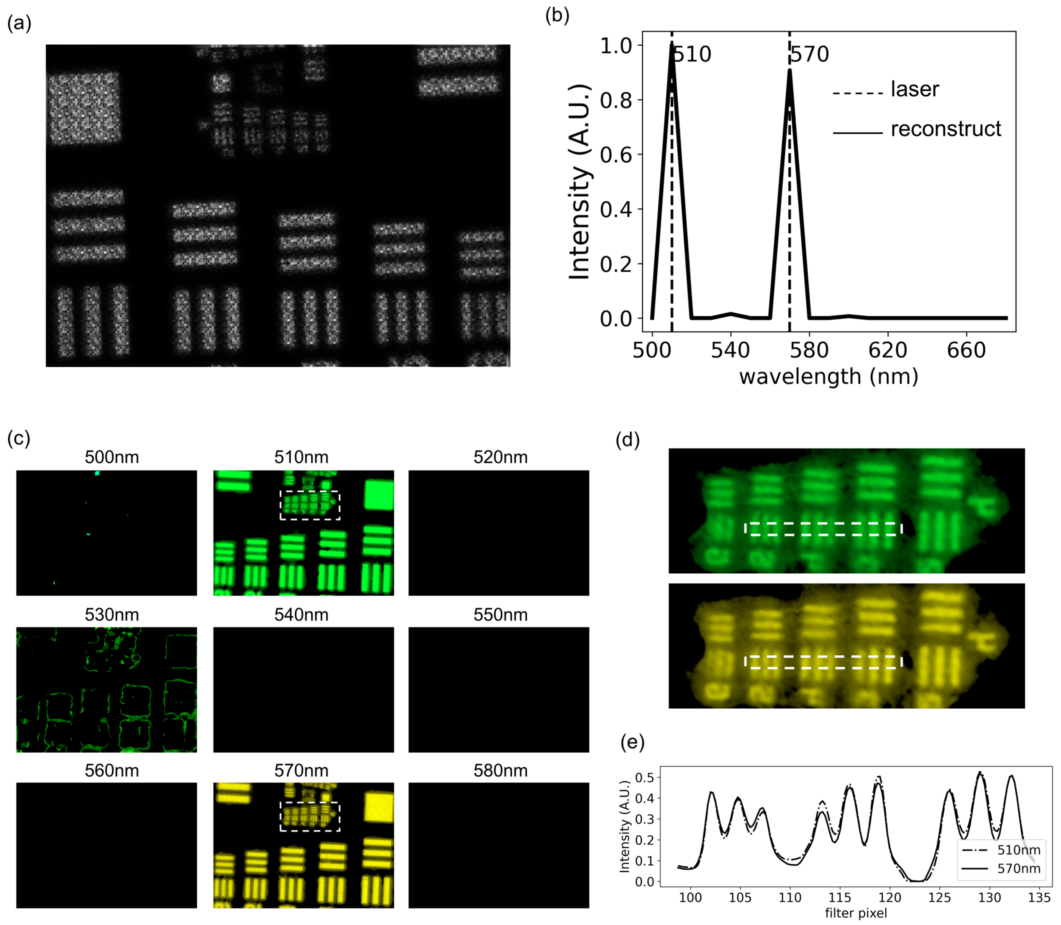

FIGURE 8.10: **(a)** The greyscale compressed image of the USAF target captured by the hyperspectral imaging system, under laser illumination at 510 nm and 570 nm. **(b)** The spectral intensity distribution of the laser output and the hyperspectral image reconstructed from the compressed image in (a). **(c)** The reconstructed hyperspectral cube in the spectral bands from 500 nm to 580 nm. **(d)** The zoom-in view of the region enclosed by the white dash line squares in (c) in the 510 nm (green) and 570 nm (yellow) spectral bands. **(e)** The intensities across the grating (along the white dashed line stripe

8.6.2 Imaging color checker

To evaluate the spectral accuracy of our hyperspectral imaging system, we imaged a Macbeth color checker and compared the reconstructed hyperspectral images with ground truth data. As shown in Figure 8.11a, broadband light (Thorlabs, SLS201/M, output wavelength range: 300 nm-2600 nm) passes through a diffuser before illuminating the color checker sample. The diffuse reflected light from the color checker is then captured by the hyperspectral imaging system described in Figure 8.11a. Due to the sharp drop in transmission of our bandpass filter at the edge of the passband

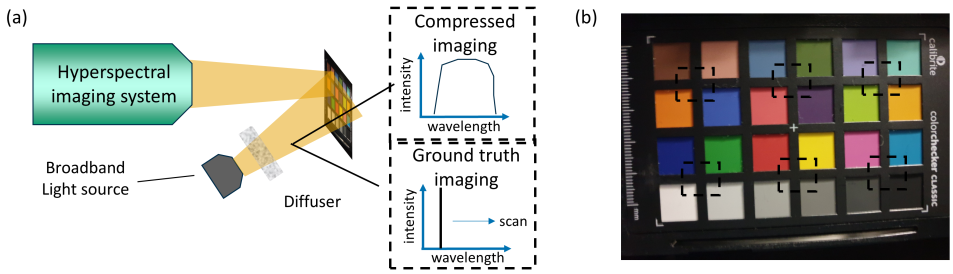

FIGURE 8.11: **(a)** Experiment setup for hyperspectral imaging of the color-chart sample under broadband ambient light. **(b)** Cellphone image of the color-chart sample. The six regions enclosed by the black dash-line are the regions being imaged.

— falling below 20% for wavelengths shorter than 500 nm or longer than 650 nm — the signal-to-noise ratio is significantly reduced at these spectral edges. Therefore, we restricted the comparison range to 500 nm–650 nm.

To acquire the ground truth hyperspectral images, we replaced the broadband light source used in Figure 8.11a with a tunable laser and removed the spectral filter array chip from the hyperspectral imaging system. We scanned the laser output across the 500 nm–650 nm range in 10 nm increments. For each wavelength, we captured an unfiltered greyscale image, which served as the ground truth hyperspectral image at that spectral band.

We captured hyperspectral images of six distinct regions of the color checker, each containing four color blocks, as shown in Figure 8.11b. Since the spectral intensity of the broadband light source is non-uniform, and the output power of the tunable laser varies with wavelength during the ground truth measurements, direct comparison between the reconstructed and ground truth images would be biased. To ensure a fair comparison, we performed white balancing by normalizing the image intensity in each wavelength band using the average intensity of the bottom-left white reference block (Block 0):

$$I_{\text{norm}}(\lambda_i, x, y) = \overline{I_{\text{ref}}(\lambda_i)} \tag{8.12}$$

where I_{norm} is the normalized image, and $\overline{I_{\text{ref}}}$ is the average intensity of the white reference block.

White balancing was applied to both the reconstructed hyperspectral images and the ground truth using the same reference block. The RGB renderings of the normalized images from all six regions of the color checker are shown in Figure 8.12a. As observed, the reconstructed images closely resemble the ground truth in overall color appearance. The colors differ from those shown in Figure 8.11b due to the restricted operating wavelength range of 500 nm–650 nm, which excludes the blue region of the spectrum. Figure 8.12b presents the spectral intensities of 24 color blocks from the reconstructed hyperspectral images in comparison with the ground truth.

We employed spectral angle mapping as an error metric to quantitatively evaluate the spectral accuracy of the reconstructed hyperspectral images. The spectral angle is computed by representing the spectra in N spectral bands as two N-dimensional vectors and calculating the angle between them. A smaller spectral angle indicates a closer match between the reconstructed spectrum and the ground truth. The spectral angles for all color blocks are displayed in Figure 8.12b. As observed, most blocks exhibit spectral angles in the range of 5° - 20°, with an average spectral angle of

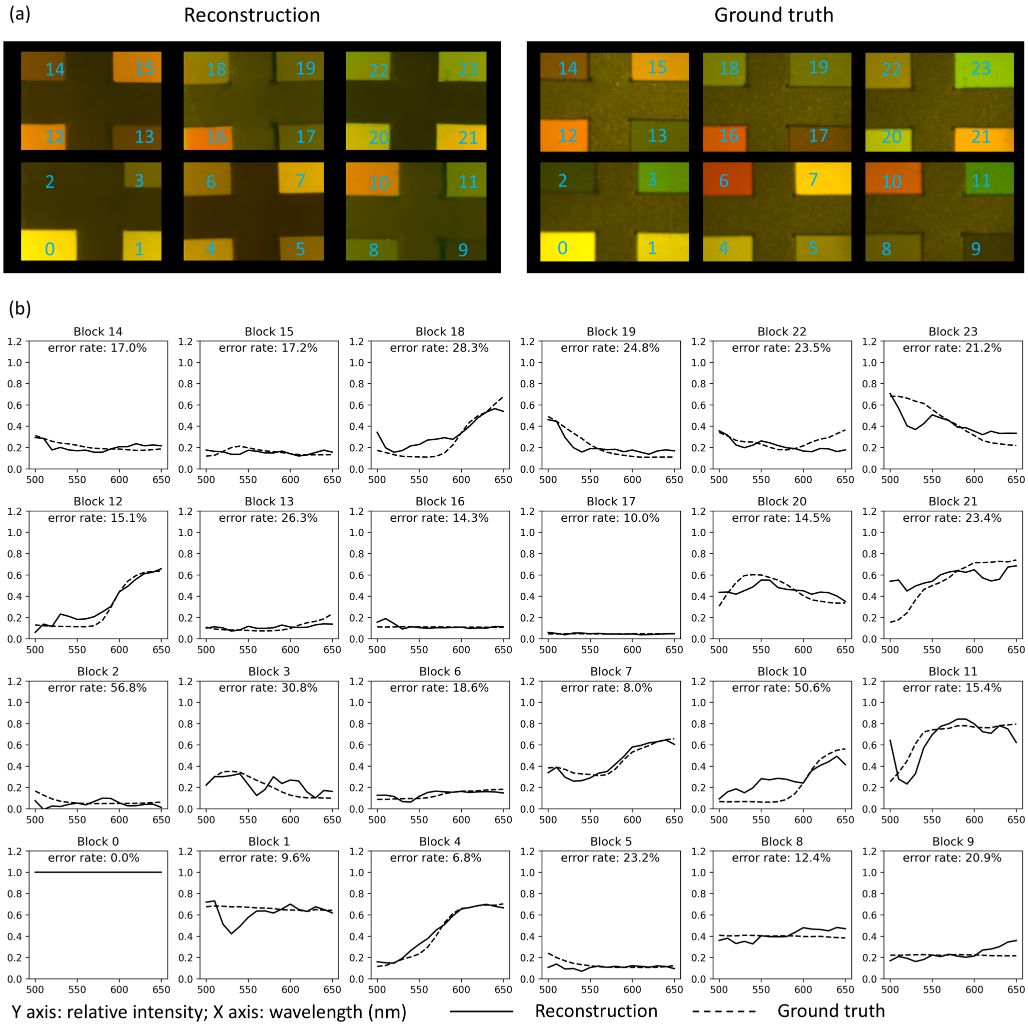

FIGURE 8.12: **(a)** The RGB rendering of reconstructed and ground truth hyperspectral images of the color-chart sample. **(b)** The spectral intensity of the 24 blocks indicated in (a), ground truth (dash line) vs. reconstruction (solid line).

14.3°. A few blocks show relatively large spectral angles, which may be attributed to the low signal-to-noise ratio caused by weak spectral intensities in the ground truth measurements for those blocks.

8.6.3 Imaging plant leaves

In addition, we performed hyperspectral imaging on natural plant leaves. The leaves were placed on a white board and covered with a glass slide. We acquire the images using the same procedure as for the color checker. White balancing was applied by normalizing the intensity at each wavelength band with the average intensity of the four corners of the image, corresponding to the white board region.

Figure 8.13a presents the RGB renderings of both the normalized ground truth and the reconstructed images of three small plant leaf samples with different colors. The reconstructed image exhibits leaf colors that closely resemble those in the ground truth. The white board appears yellow due to the limited spectral range of the hyperspectral images (500 nm–650 nm), which excludes the blue portion of the spectrum.

For each leaf, a small region (indicated in Figure 8.13a) was selected for spectral sampling. The resulting spectra are shown in Figure 8.13b. As observed, the reconstructed spectra align well with the ground truth, except for a noticeable dip at 550 nm, indicating a reconstruction artifact. This artifact likely originates from inaccuracies in the assumed pre-measurement spectral model of the leaf tissue. Incorporating stronger penalty to abrupt variations in the spectrum may help suppress the artifact.

Figure 8.13d displays the full hyperspectral image of the plant leaves across all spectral bands. The reconstructed and ground truth images show consistent spatial patterns across the spectrum, supporting the effectiveness of the reconstruction. The three leaves exhibit varying relative brightness at different wavelengths, which explains the differences in their perceived color.

To quantitatively evaluate the reconstruction accuracy of the hyperspectral image, we further calculate the spectral angle map between the ground truth and the reconstruction and plot it in Figure 8.13c. As can be seen, most of the areas have low spectral angles of $5^\circ - 10^\circ$ despite several speckles with a high spectral angle of $25^\circ - 40^\circ$. On average, the regions of the three leaves have a spectral angle of 9.7°, with a $21\times$ reduction in the number of measurements owing to our snapshot hyperspectral camera design.

8.7 Discussion

With metasurface-enabled spectral filtering combined with compressed sensing, we have demonstrated snapshot hyperspectral imaging without sacrificing spatial resolution. Compared to spectral filters based on optical absorption materials or optical cavities, metasurface-based filters offer unique advantages, including simplified fabrication, straightforward integration with camera sensors, and greater tunability that allows for design optimization.

Overall, this compressed sensing approach enables snapshot hyperspectral imaging. In contrast to conventional hyperspectral imaging strategies that rely on sequential scanning in either the spectral or spatial domain, this approach significantly reduces the acquisition time required to capture a complete hyperspectral dataset. This improvement comes at the expense of spectral accuracy, but the trade-off is acceptable for many practical applications due to the inherent signal sparsities of hyperspectral data in natural scenes.

Although snapshot imaging avoids sequential scanning, it still requires post-acquisition reconstruction, which comsumes time and demands computational resources. Nevertheless, the reconstruction algorithms are sufficiently efficient to support video-rate imaging, making this approach feasible for real-time applications such as agricultural monitoring and environmental sensing. The true advantages of snapshot hyperspectral imaging lies in scenarios where rapid image capture is essential, and extended post-processing is permissible, such as in analyzing thermal diffusion dynamics or monitoring chemical and biological reaction processes. Another important advantage arises in low-signal or photon-sensitive settings, where long integration times or minimal illumination doses are required. In these cases, snapshot acquisition is inherently more favorable than scanning-based methods.

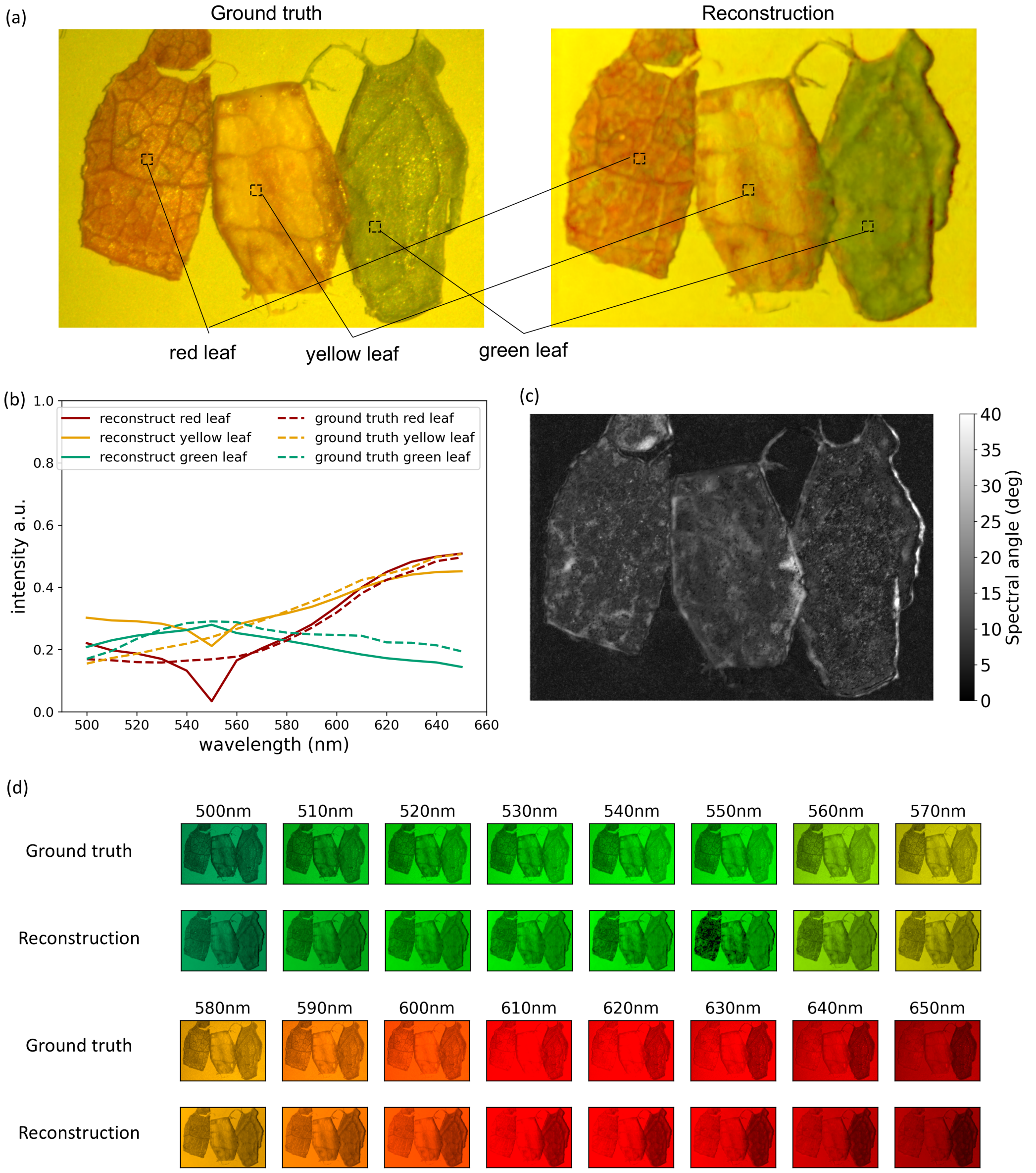

FIGURE 8.13: **(a)** The RGB rendering of the ground truth and reconstructed hyperspectral images of three plant leaves with different colors. **(b)** The ground truth and reconstructed spectra of three regions in different leaves. The sampling regions are indicated by the black squares in (a). **(c)** The spectral angle map between ground truth and reconstructed hyperspectral image. **(d)** The full hyperspectral image data of the plant leaves, ground truth vs. reconstruction.

Chapter 9

Future work

9.1 Overview

This chapter outlines some of the remaining challenges encountered in the research projects presented in this thesis and conceptually introduces ideas of potential approaches to address them. The aim is to provide inspiration and direction for future work that may build upon and extend the current research.

9.2 Integration of metalens with scanning fiber endoscope

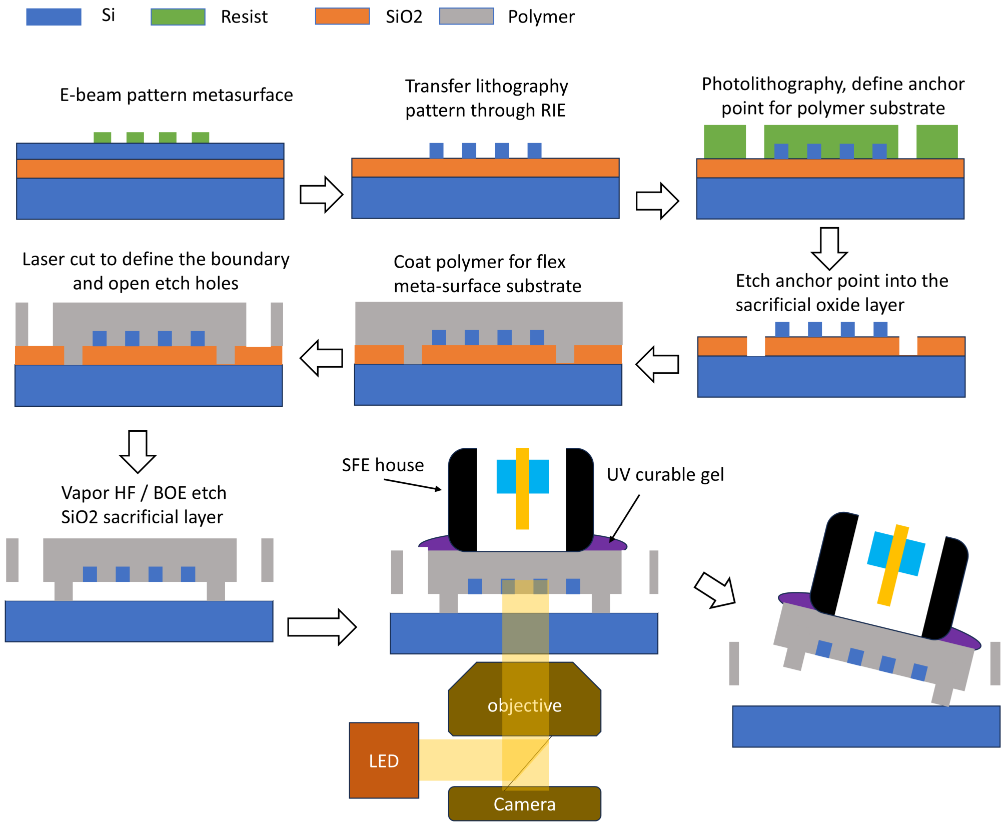

FIGURE 9.1: Proposed process flow of integrating the metalens with scanning fiber endoscope

In Chapters 5 and 6, we present the design of metalenses for scanning fiber endoscopes (SFEs). In these demonstrations, the metalenses were aligned with the scanning fiber on an optical table. Realizing a fully integrated metalens-based SFE, however, requires incorporating the metalens directly into the SFE's scanning housing—a step yet to be achieved. As shown in Chapters 5.3, metalenses are typically fabricated on rigid, thick dielectric substrates. These substrates must be at least hundreds of micrometers thick to provide the mechanical strength necessary for fabrication, but this adds significant thickness to the otherwise micrometer thin metalens. Moreover, dielectric substrates are difficult to cut, posing challenges for large-scale production and for integrating metalenses specifically designed for SFEs.

A promising strategy to overcome these limitations is to encapsulate the metasurface in a thin, flexible polymer layer [53], which can be easily cut and transferred onto a fiber [54]. However, in reported processes, the metalens is manually peeled from its substrate and repositioned before attachment, risking the damage of the metalens.

Here, we propose a modified approach in which encapsulation, cutting, transfer of the metalens and the alignment of the metalens with the SFE scanning house are performed in a continuous process, with the metalens remaining on its original substrate until final release, as illustrated in Figure 9.1. This method could enable the development of fully integrated metalens-based endoscopes that leverage the advantages of ultrathin, planar optics for easier integration and enhanced performance.

9.3 Image quality enhancement using a confocal configuration in a scanning fiber endoscope

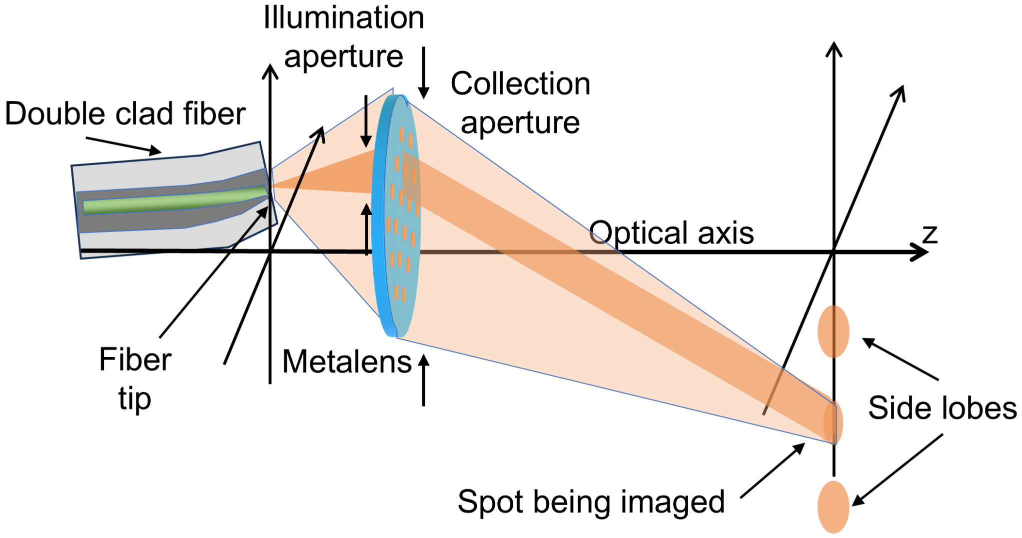

FIGURE 9.2: Proposed confocal setup for the scanning fiber endoscope.

The relatively low efficiency of the polychromatic metalens demonstrated in Chapter 6 reduces the image contrast of a multi-color endoscope, thereby degrading imaging quality. Enhancing the metalens design to increase its relative efficiency could mitigate this issue, but achieving such improvement remains a longstanding challenge. In addition to design optimization, incorporating a confocal configuration into the SFE can further improve image contrast by exploiting the spatial filtering effect of confocal optics [27, 55]. As shown in the proposed confocal SFE setup (Figure 9.2), a single spot on the sample is illuminated by the beam emitted from the scanning fiber tip and focused by the metalens during scanning. Light scattered from this spot is then re-collected by the metalens and refocused back into the same fiber core, rather than being captured by surrounding fibers as in the state-of-the-art SFE.

Due to the metalens's low efficiency, a portion of the beam energy is not concentrated on the intended focal spot. As shown by the simulations and experiments in Chapter 6, two prominent side lobes are present. In a conventional SFE, where scattered light is collected by surrounding fibers, these side lobes generate shadow artifacts and reduce image contrast (see Chapter 6.7). In contrast, the confocal configuration largely rejects light from the side lobes, as it does not couple efficiently back into the fiber core, thereby suppressing unwanted signals and improving image contrast.

In a traditional confocal arrangement, the same fiber core serves as both the light emitter and the collector. Because the fiber's numerical aperture (NA) defines both the emission and acceptance angles, only a small section of the metalens is used for both illumination and collection. This maximizes spatial filtering, yielding the highest imaging contrast, but limits the collection aperture and thus the overall light collection efficiency. Given that the permissible laser power for illuminating biological tissue is often restricted, maximizing collection efficiency is essential. This can be addressed using a double-clad fiber (Figure 9.2). In this configuration, the single-mode core with a small NA delivers the illumination beam, while the inner cladding, having a larger NA, collects scattered light over a wider acceptance angle—effectively from the entire metalens. This increases light collection efficiency at the expense of partially reduced spatial filtering. Additionally, because light from the entire metalens suffers from more aberrations than those from only a small section of the metalens, the collected light forms an expanded spot on the fiber's scanning plane. Therefore, the inner cladding diameter must be designed to accommodate this spot size, preserving effective signal collection.

9.4 Far-field and large number of pixel spectrally encoding imaging

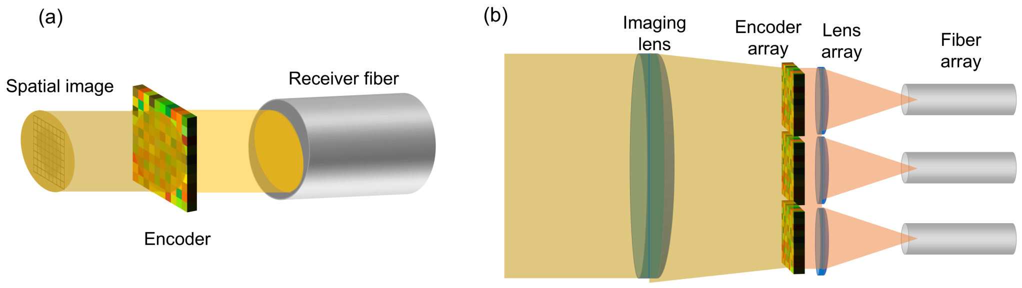

FIGURE 9.3: **(a)** Single fiber, near-field spectral endocing imaging. **(b)** Proposed configuration for massive parallel far-field spectrally encoding imaging with fiber array.

The spectrally encoding imaging demonstrated in Chapter 7 operates in the near field: the object must be placed in close proximity to the encoder, and the light-collecting fiber is also positioned near the encoder, as shown in Figure 9.3a. This requirement limits its practical use in endoscopic applications, since the endoscope tip cannot always be placed directly adjacent to the target tissue. Furthermore, as a proof-of-concept, we demonstrated imaging through only a single fiber, achieving merely 16 pixels—an insufficient number for most practical uses.

To develop a spectrally encoding endoscope with higher pixel counts capable of doing far-field imaging, and to overcome the pixel-density limitation (i.e., the maximum number of pixels that can be realized in a given cross-sectional area of the endoscope tip, as discussed in Chapter 3.2.3), a natural approach is to employ a dense fiber array, such as a coherent fiber bundle, combined with an array of spectral encoders and an imaging lens, as illustrated in Figure 9.3b. In this configuration, the encoder array is placed at the image plane of the lens, where it encodes the spatial information into spectra. Each encoder in the array is paired with a corresponding fiber, with a microlens coupling the encoded light into the fiber. While the physical constraints of fiber optics require core-to-core spacing on the order of several micrometers to prevent cross-talk between each core, the spectral encoder itself can incorporate filter blocks with dimensions as small as one micrometer or even sub-micrometer. Consequently, each encoder can transfer multiple spatial pixels through a single fiber by encoding them spectrally. This configuration enables a spectrally encoding endoscope to achieve a total pixel number far exceeding that of a conventional coherent fiber bundle endoscope, where each fiber core transmits only one spatial pixel.

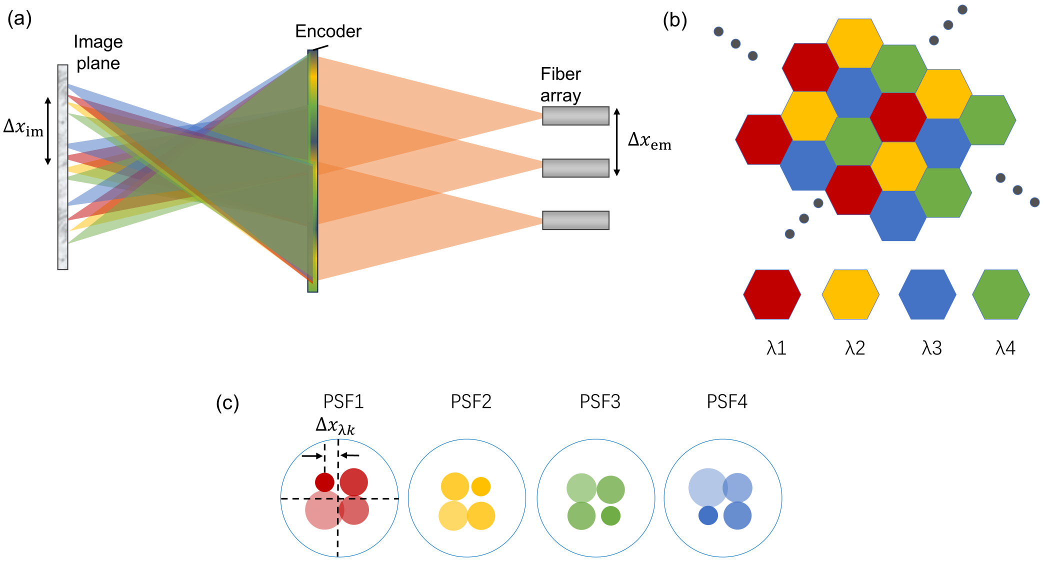

FIGURE 9.4: **(a)** Proposed configuration for far-field spectrally encoding imaging with an encoding lens. **(b)** Demonstration of spatial division multiplexing design of the encoding lens. **(c)** Expected point-spread-function of the encoding lens at different wavelengths.

Such a configuration requires complex integration and precise alignment between the fiber array and the encoder array, which poses significant fabrication challenges and reduces the robustness of the endoscopic system, since even small misalignments can cause substantial performance degradation. This hinders the commercialization of spectrally encoded endoscopes. To address this limitation, we propose an alternative solution for far-field, high-pixel-count spectrally encoding imaging, illustrated in Figure 9.4a, which eliminates the need for precise alignment between the encoder and the fiber array.

This setup is adapted from a conventional confocal imager: light emitted from fibers located at positions x_{em} is projected by a lens onto the image plane at positions x_{im}, scattered back, and then re-collected by the same lens and coupled into the originating fibers. In this way, the fiber core at x_{em} records the image information at x_{im}. The mapping $x_{\text{em}} \rightarrow x_{\text{im}}$ follows the standard imaging law of a lens. The fiber array has a core-to-core spacing of Δx_{em}, which defines the pixel pitch Δx_{im} on the image plane (Figure 9.4a). Typically, Δx_{em} is larger than the size of the beam spot in the image plane, which causes the system not to reach the diffraction limit.

In a spectrally encoding endoscope, the lens is modified to function as a spectral encoder by introducing unconventional dispersion. Specifically, for light at wavelength λ_k, the focal spot on the image plane is displaced from x_{im} by $\Delta x_{\lambda k}$, with $\Delta x_{\lambda k} \in (0, \Delta x_{\text{im}})$. These wavelength-dependent focal spot shifts encode fine spatial details near x_{im} into the spectrum collected by the fiber at x_{em}. As a result, each fiber captures a low-resolution spectral image around x_{im}, and the collection of these images across the fiber array can be stitched into a large image. The total number of resolvable pixels is given by the number of fiber cores multiplied by the number of wavelength bands, thereby exceeding the pixel-density limitation discussed in Chapter 3.2.3.

This wavelength-dependent vari-focal lens can be realized through spatial division multiplexing. The metalens is divided into n regions, where n is the number of wavelength bands. Each region, indexed by k, is designed to focus light at λ_k onto a lateral position shifted by $\Delta x_{\lambda k}$. Figure 9.4b shows an example of a four-wavelength varifocal metalens, where the red, yellow, green, and blue regions are assigned to λ_1, λ_2, λ_3, and λ_4, respectively, and the corresponding focal spots are positioned at the top-left, top-right, bottom-right, and bottom-left of the center. Similar spatial multiplexing strategies have been employed in the development of metalens spectrometers [56, 57].

In this spatial multiplexing approach, the on-focus spot for one wavelength will result in defocused spots at other wavelengths. Thus, for a four-wavelength vari-focal lens, at each λ_k, the point spread function PSF_{λ_k} consists of one on-focus spot and three defocused spots. However, the locations of these spots differ for each wavelength (Figure 9.4), resulting in wavelength-dependent PSFs that encode spatial information into the spectrum. The inverse design methodology introduced in Chapter 6.3 can be applied to reduce the correlation between PSF_{λ} across different λ, thereby enhancing the robustness of spatial information reconstruction from the spectrum.

Bibliography

[1] F. Litvack et al. "Angioscopic Visualization of Blood-Vessel Interior in Animals and Humans". In: *Clinical Cardiology* 8.2 (1985). Abf96 Times Cited:51 Cited References Count:14, pp. 65–70. DOI: `DOI10.1002/clc.4960080202`.

[2] L. E. Savastano et al. "Diagnostic and Interventional Optical Angioscopy in Ex Vivo Carotid Arteries". In: *Operative Neurosurgery* 13.1 (2017), pp. 36–46. DOI: `10.1093/ons/opw002`.

[3] P. Z. McVeigh et al. "High-Resolution Scanning Fiber Angioscopy as an Adjuvant to Fluoroscopy During Endovascular Interventions". In: *Journal of Endovascular Therapy* 25.5 (2018), pp. 617–623. DOI: `10.1177/1526602818794663`.

[4] K. W. Shim et al. "Neuroendoscopy : Current and Future Perspectives". In: *Journal of Korean Neurosurgical Society* 60.3 (2017), pp. 322–326. DOI: `10.3340/jkns.2017.0202.006`.

[5] H. Arishima et al. "Spinal endoscopy combined with selective CT myelography for dural closure of the spinal dural defect with superficial siderosis: technical note". In: *Journal of Neurosurgery-Spine* 28.1 (2018), pp. 96–102. DOI: `10.3171/2017.5.Spine17233`.

[6] Y. Liu et al. "Meta-objective with sub-micrometer resolution for microendoscopes". In: *Photonics Research* 9.2 (2021), pp. 106–115. DOI: `10.1364/Prj.406197`.

[7] J. E. Fröch et al. "Real time full-color imaging in a Meta-optical fiber endoscope". In: *Elight* 3.1 (2023). DOI: `ARTN1310.1186/s43593-023-00044-4`.

[8] X. Ye et al. "Chip-scale metalens microscope for wide-field and depth-of-field imaging". In: *Advanced Photonics* 4.4 (2022). DOI: `10.1117/1.Ap.4.4.046006`.

[9] H. Pahlevaninezhad et al. "Nano-optic endoscope for high-resolution optical coherence tomography in vivo". In: *Nature Photonics* 12.9 (2018), pp. 540–+. DOI: `10.1038/s41566-018-0224-2`.

[10] C. M. Lee et al. "Scanning fiber endoscopy with highly flexible, 1 mm catheterscopes for wide-field, full-color imaging". In: *Journal of Biophotonics* 3.5-6 (2010), pp. 385–407. DOI: `10.1002/jbio.200900087`.

[11] R. Barankov and J. Mertz. "High-throughput imaging of self-luminous objects through a single optical fibre". In: *Nature Communications* 5 (2014). DOI: `ARTN558110.1038/ncomms6581`.

[12] S. M. Kolenderska et al. "Scanning-free imaging through a single fiber by random spatio-spectral encoding". In: *Optics Letters* 40.4 (2015). Cb7zu, pp. 534–537. DOI: `10.1364/Ol.40.000534`.

[13] A. Rajiv et al. "Electromechanical Model-Based Design and Testing of Fiber Scanners for Endoscopy". In: *Journal of Medical Devices-Transactions of the Asme* 12.4 (2018). DOI: `Artn04100310.1115/1.4040271`.

[14] Ningzhi Xie et al. "Large field-of-view short-wave infrared metalens for scanning fiber endoscopy". In: *Journal of Biomedical Optics* 28.09 (Mar. 2023). DOI: `10.1117/1.jbo.28.9.094802`.

[15] Victor Liu and Shanhui Fan. "S4 : A free electromagnetic solver for layered periodic structures". In: *Computer Physics Communications* 183.10 (2012), pp. 2233–2244. DOI: `https://doi.org/10.1016/j.cpc.2012.04.026`.

[16] Ningzhi Xie et al. "Inverse-designed large field-of-view polychromatic metalens for tri-color scanning fiber endoscopy". In: *Communications Engineering* 4.1 (Mar. 2025). DOI: `10.1038/s44172-025-00377-7`.

[17] L. C. Huang et al. "Full-Color Metaoptical Imaging in Visible Light". In: *Advanced Photonics Research* 3.5 (2022). DOI: `ARTN210026510.1002/adpr.202100265`.

[18] W. T. Chen et al. "A broadband achromatic metalens for focusing and imaging in the visible". In: *Nature Nanotechnology* 13.3 (2018), pp. 220–226. DOI: `10.1038/s41565-017-0034-6`.

[19] S. M. Wang et al. "A broadband achromatic metalens in the visible". In: *Nature Nanotechnology* 13.3 (2018), pp. 227–232. DOI: `10.1038/s41565-017-0052-4`.

[20] Zhi-Bin Fan et al. "A broadband achromatic metalens array for integral imaging in the visible". en. In: *Light Sci. Appl.* 8.1 (July 2019), p. 67.

[21] Ren Jie Lin et al. "Achromatic metalens array for full-colour light-field imaging". en. In: *Nat. Nanotechnol.* 14.3 (Mar. 2019), pp. 227–231.

[22] Haoran Ren et al. "An achromatic metafiber for focusing and imaging across the entire telecommunication range". en. In: *Nat. Commun.* 13.1 (July 2022), p. 4183.

[23] H. R. Ren et al. "An achromatic metafiber for focusing and imaging across the entire telecommunication range". In: *Nature Communications* 13.1 (2022). DOI: `ARTN418310.1038/s41467-022-31902-3`.

[24] L. C. Huang et al. "Design and analysis of extended depth of focus metalenses for achromatic computational imaging". In: *Photonics Research* 8.10 (2020), pp. 1613–1623. DOI: `10.1364/Prj.396839`.

[25] E. Arbabi et al. "Multiwavelength polarization-insensitive lenses based on dielectric metasurfaces with meta-molecules". In: *Optica* 3.6 (2016), pp. 628–633. DOI: `10.1364/Optica.3.000628`.

[26] Luocheng Huang et al. "Broadband thermal imaging using meta-optics". en. In: *Nat. Commun.* 15.1 (Feb. 2024), p. 1662.

[27] E. S. Barhoum, R. S. Johnston, and E. J. Seibel. "Optical modeling of an ultrathin scanning fiber endoscope, a preliminary study of confocal versus non-confocal detection". In: *Optics Express* 13.19 (2005), pp. 7548–7562. DOI: `Doi10.1364/Opex.13.007548`.

[28] D. Yelin et al. "Double-clad fiber for endoscopy". In: *Optics Letters* 29.20 (2004), pp. 2408–2410. DOI: `Doi10.1364/Ol.29.002408`.

[29] Ningzhi Xie et al. "Spectrally Encoded Nonscanning Imaging through a Fiber". In: *ACS Photonics* 11.3 (Feb. 2024), 1117–1124. DOI: `10.1021/acsphotonics.3c01582`.

[30] Yu Horie et al. "Wide bandwidth and high resolution planar filter array based on DBR-metasurface-DBR structures". In: *Optics Express* 24.11 (2016), pp. 11677–11682. DOI: `10.1364/oe.24.011677`.

[31] Eric Swanson. "Multicore Fiber Instrument with 3D-Printed Distal Optics". Pat. US20210149101A1. 2021.

[32] Z. Wang et al. "Single-shot on-chip spectral sensors based on photonic crystal slabs". In: *Nature Communications* 10 (2019). DOI: `ARTN102010.1038/s41467-019-08994-5`.

[33] Lankapalli Ravikanth et al. "Extraction of Spectral Information from Hyperspectral Data and Application of Hyperspectral Imaging for Food and Agricultural Products". In: *Food and Bioprocess Technology* 10.1 (Nov. 2016), 1–33. DOI: `10.1007/s11947-016-1817-8`.

[34] Peter Favreau et al. "Thin-film tunable filters for hyperspectral fluorescence microscopy". In: *Journal of Biomedical Optics* 19.1 (Sept. 2013), p. 011017. DOI: `10.1117/1.jbo.19.1.011017`.

[35] Chen Zhang et al. "A novel 3D multispectral vision system based on filter wheel cameras". In: *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, Oct. 2016, 267–272. DOI: `10.1109/ist.2016.7738235`.

[36] Samir Sahli et al. "Hyperspectral imaging: comparison of acousto-optic and liquid crystal tunable filters". In: *Medical Imaging 2018: Physics of Medical Imaging*. Ed. by Guang-Hong Chen, Joseph Y. Lo, and Taly Gilat Schmidt. SPIE, Mar. 2018, p. 96. DOI: `10.1117/12.2282532`.

[37] Jianwei Qin et al. "Line-Scan Hyperspectral Imaging Techniques for Food Safety and Quality Applications". In: *Applied Sciences* 7.2 (Jan. 2017), p. 125. DOI: `10.3390/app7020125`.

[38] Jiaqiong Li et al. "4D line-scan hyperspectral imaging". In: *Optics Express* 29.21 (Oct. 2021), p. 34835. DOI: `10.1364/oe.441213`.

[39] Joaquim J. Sousa et al. "UAV-Based Hyperspectral Monitoring Using Push-Broom and Snapshot Sensors: A Multisite Assessment for Precision Viticulture Applications". In: *Sensors* 22.17 (Aug. 2022), p. 6574. DOI: `10.3390/s22176574`.

[40] Bert Geelen, Nicolaas Tack, and Andy Lambrechts. "A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic". In: *Advanced Fabrication Technologies for Micro/Nano Optics and Photonics VII*. Ed. by Georg von Freymann, Winston V. Schoenfeld, and Raymond C. Rumpf. Vol. 8974. SPIE, Mar. 2014, p. 89740L. DOI: `10.1117/12.2037607`.

[41] Yuri Murakami, Masahiro Yamaguchi, and Nagaaki Ohyama. "Hybrid-resolution multispectral imaging using color filter array". In: *Optics Express* 20.7 (Mar. 2012), p. 7173. DOI: `10.1364/oe.20.007173`.

[42] Jian Zhang et al. "Image Compressive Sensing Recovery via Collaborative Sparsity". In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 2.3 (Sept. 2012), 380–391. DOI: `10.1109/jetcas.2012.2220391`.

[43] Meenu Rani, S. B. Dhok, and R. B. Deshmukh. "A Systematic Review of Compressive Sensing: Concepts, Implementations and Applications". In: *IEEE Access* 6 (2018), 4875–4894. DOI: `10.1109/access.2018.2793851`.

[44] Shuyuan Yang et al. "Compressive Hyperspectral Imaging via Sparse Tensor and Nonlinear Compressed Sensing". In: *IEEE Transactions on Geoscience and Remote Sensing* 53.11 (2015), pp. 5943–5957. DOI: `10.1109/TGRS.2015.2429146`.

[45] Motoki Yako et al. "Video-rate hyperspectral camera based on a CMOS-compatible random array of Fabry–Pérot filters". In: *Nature Photonics* 17.3 (Jan. 2023), 218–223. DOI: `10.1038/s41566-022-01141-5`.

[46] Liheng Bian et al. "A broadband hyperspectral image sensor with high spatio-temporal resolution". In: *Nature* 635.8037 (Nov. 2024), 73–81. DOI: `10.1038/s41586-024-08109-1`.

[47] Lidan Zhang et al. "Real-time machine learning–enhanced hyperspectro-polarimetric imaging via an encoding metasurface". In: *Science Advances* 10.36 (Sept. 2024). DOI: `10.1126/sciadv.adp5192`.

[48] Jussi PS Parkkinen, Jarmo Hallikainen, and Timo Jaaskelainen. "Characteristic spectra of Munsell colors". In: *Journal of the Optical society of America A* 6.2 (1989), pp. 318–322.

[49] Vishwanath Saragadam et al. “DeepTensor: Low-Rank Tensor Decomposition With Deep Network Priors”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46.12 (2024), pp. 10337–10348.

[50] Vishwanath Saragadam et al. “Wire: Wavelet implicit neural representations”. In: *Computer Vision and Pattern Recognition*. 2023.

[51] Kazuma Shinoda, Yasuo Ohtera, and Madoka Hasegawa. “Snapshot multispectral polarization imaging using a photonic crystal filter array”. In: *Optics Express* 26.12 (June 2018), p. 15948. DOI: `10.1364/oe.26.015948`.

[52] Zhu Wang et al. “Single-shot on-chip spectral sensors based on photonic crystal slabs”. In: *Nature Communications* 10.1 (Mar. 2019). DOI: `10.1038/s41467-019-08994-5`.

[53] S. M. Kamali et al. “Decoupling optical function and geometrical form using conformal flexible dielectric metasurfaces”. In: *Nature Communications* 7 (2016). DOI: `ARTN1161810.1038/ncomms11618`.

[54] Fei He et al. “Diffractive metasurface light-shaping from fiber endoscope probes for increased depth of field”. In: *Endoscopic Microscopy XVIII*. Ed. by Guillermo J. Tearney M.D., Thomas D. Wang, and Melissa J. Suter. Vol. 12356. International Society for Optics and Photonics. SPIE, 2023, p. 1235606. DOI: `10.1117/12.2648860`.

[55] Eric J. Seibel. “Medical imaging, diagnosis, and therapy using a scanning, single optical fiber system”. Patent US6975898B2. U.S. Patent No. 6,975,898. 2005.

[56] Feng Tang et al. “Dielectric metalenses at long-wave infrared wavelengths: Multiplexing and spectroscope”. In: *Results in Physics* 18 (Sept. 2020), p. 103215. DOI: `10.1016/j.rinp.2020.103215`.

[57] Ruoxing Wang et al. “Compact multi-foci metalens spectrometer”. In: *Light: Science amp; Applications* 12.1 (May 2023). DOI: `10.1038/s41377-023-01148-9`.