Novel Devices for Dynamically Shaping the Wavefront and Polarization of Light

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>She, Alan Jenting. 2018. Novel Devices for Dynamically Shaping the Wavefront and Polarization of Light. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:42015658">http://nrs.harvard.edu/urn-3:HUL.InstRepos:42015658</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Novel Devices for Dynamically Shaping the Wavefront and Polarization of Light

A dissertation presented by Alan Jenting She to The John A. Paulson School of Engineering and Applied Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the subject of Applied Physics at Harvard University Cambridge, Massachusetts December 2017
Novel Devices for Dynamically Shaping the Wavefront and Polarization of Light

Abstract

Light, which provides the human eye with the sense of vision, is the basis for many key technologies used in everyday life, and the advanced control of which also benefits technological progress in other areas. The wavefront and polarization are two important properties of light, or, in general, electromagnetic radiation, that dictate its behavior. The ability to precisely control these properties as a function of time is highly desirable, as it is essential for numerous powerful applications, including imaging, high tech manufacturing, and communications, to name a few. There has been a bulk of work on the manipulation of these properties, but there remains much to be improved in terms of extreme compactification, reduction of power, and increase in speed and precision, as use cases are driven towards untethered, lightweight, and high performance devices. In
this thesis, novel devices are presented that dynamically shape wavefront and polarization of light, using new principles of operation and design methods.

Wavefront shaping is achieved using metasurfaces, in which, recently, there has been significant scientific and technological interest due to their inherent multifunctionality and potential to dramatically reduce the thickness of the optics in a wide range of applications. The challenge, however, in making practical devices with metasurfaces is the difficulty in fabricating them with large areas. One of the key obstacles in this endeavor is the enormous data density required by the subwavelength resolution criterion imposed over large areas, resulting in giant file sizes for the layout design files describing structures greater than a few hundred microns in diameter. We present a scalable metasurface layout compression algorithm that exponentially reduce the design file size of (by 3 orders of magnitude for a centimeter diameter metalens, and even greater gains in compression for larger sizes) and a route to mass manufacturing of metasurface lenses (metalenses) using stepper photolithography with extremely large areas, up to (but not limited to) centimeters in diameter. Because of what is demonstrated here, the claim that chipmakers will be making lenses of the future (as well as the chips), i.e. the unification of two industries: semiconductor manufacturing and lens-making, is now closer to reality. However, metasurface devices are by themselves static, such that additional steps are required in order to introduce the feature of tunability, such as in focal length or magnification control. To take advantage of the thinness and planarity of metalenses, electrical tuning of lateral motion is essential for focal length and magnification control, which in conventional optical systems is performed by longitudinal mechanical motion along the optical axis. A large area, integrated optical
device, no more than 30 microns thick, is presented, which imprints a strain field onto the optical wavefront, by way of a soft metasurface intermediary, enabling simultaneous control over focal length, astigmatism, and image shift. These combined capabilities have so far been possible only in electron optics, and has been made possible by combining metalenses, artificial muscles (i.e., dielectric elastomer actuators (DEAs)), and carbon nanotube-based stretchable transparent electrodes. This is the first step of many that seeks to exert arbitrary control over the strain profile of a metasurface: transforming a transformation optic.

Polarization shaping is achieved using a new architecture. Previously, the general problem of generating arbitrary time-varying states of polarization (SOP) has always been mathematically formulated by a series of linear transformations, i.e. a product of matrices, imposing a serial architecture, such as a series of rotating wave plates. An alternative parallel architecture is presented, which is described by a sum, rather than a product, of matrices. The theory is experimentally demonstrated by modulating spatially-separated polarization components of a laser using a digital micromirror device (DMD) that are subsequently beam combined. This method greatly expands the parameter space for engineering devices that control polarization. Performance characteristics, such as speed, stability, and spectral range, are entirely dictated by the technologies of optical intensity modulation – absorption, reflection, emission, and scattering, any of which may be used.

Our results have the potential of opening up a wealth of future new applications, including electrical control over all major optical aberrations and polarization, which could lead to major advances in optical microscopes and imaging systems as well as
portable and wearable devices. Our results demonstrate the possibility of future optical microscopes, which operate fully electronically, as well as compact optical systems that rapidly probe various polarization states or employ the principles of adaptive optics to correct many orders of aberrations, simultaneously.
# Table of Contents

Abstract ................................................................................................................................. iii
Table of Contents .................................................................................................................. vii
List of Figures ........................................................................................................................ ix
Acknowledgments ................................................................................................................ xx

Chapter 1 Introduction .......................................................................................................... 1
  1.1 Properties of light ............................................................................................................. 1
  1.2 Shaping the wavefront of light ....................................................................................... 4
  1.3 Metasurfaces .................................................................................................................. 7
  1.4 Shaping the polarization of light .................................................................................. 10
  1.5 Summary ....................................................................................................................... 12

Chapter 2 Scalable enlargement of metasurfaces for practical purposes ..................... 13
  2.1 Introduction .................................................................................................................... 13
  2.2 METAC algorithm ......................................................................................................... 14
  2.3 Large area lenses: fabrication and experiment ............................................................. 29
  2.4 Outlook: lenses of the future ......................................................................................... 37

Chapter 3 Dynamic wavefront shaping with adaptive metalenses .................................. 40
  3.1 Introduction .................................................................................................................... 40
  3.2 Theory of strain-field mediated tuning ......................................................................... 42
  3.3 Device design, fabrication, and assembly ..................................................................... 50
3.4 Experiment and characterization ................................................................. 66
3.5 Outlook: compact adaptive optics ............................................................... 75

Chapter 4 Dynamic polarization shaping with parallel polarization state generation .... 77

4.1 Introduction .................................................................................................... 77
4.2 Theory of parallel polarization state generation .......................................... 81
4.3 Device design and assembly ......................................................................... 92
4.4 Experiment .................................................................................................... 94
4.5 Outlook: challenges and applications ............................................................. 96

Bibliography ....................................................................................................... 101
List of Figures

Figure 1-1. Wave nature of light. Light is shown as a transverse electromagnetic wave with wavelength of $\lambda$ and propagating in space ($x$) in the direction of its wavevector, $k$. The electric and magnetic field components are shown in blue and red, respectively.

Figure 1-2 Illustration of the wavefront and polarization of an example light wave extended in two dimensions, $x$ and $y$, and polarized (red arrows) in the out of plane direction, $z$. The locus of points with the same phase (“iso-phase”) illustrates the wavefronts of light – here they are shown as black dotted lines. The direction perpendicular to the wavefront corresponds to the local propagation direction of the light wave ($k$).

Figure 1-3. Lenses as wavefront shapers. Lenses and many other optical devices, in general, can be considered as objects whose primary functions are to transform the incident wavefront into a desired output going wavefront. Here a thin lens, represented as a thin yellow line, focuses an incoming plane wave ($\phi_{in}=2\pi x/\lambda$) to a focal point (where $x_0$ is the focal length) by transforming the planar wavefront into a spherical one ($\phi_{out}=2\pi (x^2+y^2)^{1/2}/\lambda$). The lens function ($\phi_{lens}$) is then simply the phase profile that provides the phase shifts that brings $\phi_{in}$ in phase with $\phi_{out}$: $\phi_{in} + \phi_{lens} = \phi_{out} + constant$. Since $\phi_{in}$ provides only a constant phase offset in the $y$-direction of the lens, it can be neglected, thus giving: $\phi_{lens} = \phi_{out}(x_0)= 2\pi(x_0^2+y^2)^{1/2}/\lambda$.

Figure 1-4. The human eye. The human eye is an advanced imaging system that is capable of adapting the shape of lens using small muscles to focus light from objects at different distances onto the retina. Image credit: By Jmarchn (Own work) [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons.

Figure 2-1. Metasurface lens design. (a) Schematic of a metasurface lens (metalens) that is designed to focus light of normal incidence, where $D$ is diameter and $f$ is focal length. The phase profile (blue curve) is implemented by a dense array of microscopic meta-elements, made of amorphous silicon cylindrical posts, supported by a SiO$_2$ substrate. (b) The diameter of each meta-element is used to control its phase response. The phase response (blue) and transmittance (red) is plotted as a function of post diameter.

Figure 2-2. FDTD simulations of the focal spot of a 2 cm diameter metalens: (a) Simulated distribution of the electric field intensity (normalized $|E|^2$) of the focal spot. (b) The cross section of intensity profile at $Y = 0$ (white dashed line in (a)) from which the size of the focused beam can be determined. This simulation
indicates that the focal spot size, i.e. the full beam waist at $1/e^2$ of the peak intensity, is 17.1 µm. Figure 2-3. Illustration of an example of data representation of rotationally symmetrically arranged cells at a specific radial position. (a) S1 is the structure of the primitive cell in the library set, which contains the structure to be copied around the circumference of the radial position. In this case, the structure in the cell is a circle of a specified diameter. (b) S2 is the next higher level structure in the library set, which contains a reference S1 as well as a duplicate reference to S1 arranged so that it is adjacent to the first reference of S1. Thus S2 contains twice as many copies as the lower level structure. The base copy angle, α, is shown. (c) S3 is the next higher level structure in the library set, which contains two references to S2, and thus 4 copies of S1. (d) Similarly, S4 references S3 twice and contains 8 copies of S1. (e) S5 references S4 twice and contains 16 copies of S1. (f) The top level structure generates the complete device structure at the specified radial position by including the minimum number of unique library references, which includes S3, S4, and S5. Figure 2-4. Multiple levels of METAC. Illustration of a possible design of a metasurface lens, where different levels are examined to show the hierarchy of top level structure. The annular structure in Figure 2-3 is actually the 5th ring of this figure, counting from the center. The level numbers indicate which of the referenced structures in the library set are shown with at least this number of levels. For example, the 5th ring only begins to appear when level 3 is shown, corresponding to the reference to the structure S3 in the library set. Levels 4 and 5 show the completion of the 5th ring, corresponding to structures S4 and S5. Only 7 levels are required to complete this device, but in general, larger devices require more levels. Figure 2-5. High level procedure for generating layout design file. First, a radial design is determined, in which the geometry of each structure is defined at each radial position. The geometry can be any polygonal shape, and is not limited to circles. Next an empty top level structure is created to contain the final device structure. Finally a loop is performed, where the structure at each radial position is efficiently copied around the arc corresponding to the circumference at the radial position. The copy mechanism is comprised of three main steps: calculation of replication parameters, generation of the library set, and creation of references to the library set, each of which is described in Figure 2-6, Figure 2-7, Figure 2-8, respectively. Figure 2-6. Calculation of replication parameters. First, the azimuthal angular separation between cells at the same radial position is calculated, which is also referred to as the base copy angle, α. This angle can be determined by maximizing the cell density or otherwise defined. Then, α is used to calculate the total number of copies, N, including the original, for that radial position. Next, N is converted from a decimal to a binary number, $N_{bin}$, which is used to determine the references to the library set in a later step, described in Figure 2-8. With that, the calculation of replication parameters is finished for this particular radial position.
Figure 2-7. Generation of library set. First, the primitive structure, S1, is created, which contains the primitive cell, which we would like to be copied. Then a loop is performed, in which increasingly higher level structures are created, each of which contain two references to the next lower level. If the number of levels exceeds the maximum number of allowed levels, which is defined here as $L_{\text{max}} = \text{maximum number of allowed levels} - 2$, then a subroutine is activated in which a structure is created without exceeding the maximum number of allowed levels (2 additional levels are created by this subroutine and also the top level structure). For examples, if there is a limit of 16 levels, then $L_{\text{max}} = 14$. In this way, a library set containing exponentially duplicated copies can be represented.

Figure 2-8. Library-set reference creation. Flowchart shows the procedure for creating references to the library set in the top level structure. To efficiently represent the intended structure using the library set, only the minimum number of unique structures in the library set should be included. The unique structures from the library set to be used can be determined by the calculated $N_{\text{bin}}$. A loop is performed in which each digit of $N_{\text{bin}}$ is queried, such that if the digit is equal to 1, then the structure in the library set with the level number corresponding to the same digit number is referenced in the top level structure; and if the digit is equal to 0, then the reference is not made. Each reference is made at the angular position such that it is adjacent to the previous reference, if it is not the first reference. The first reference may be made at any initial angle, $\gamma$.

Figure 2-9. METAC algorithm performance. (c) Due to the extremely large number of meta-elements required to comprise a large area metalens, a metasurface data compression algorithm, which we call METAC, was developed to generate manageable file sizes of metalens designs. Four methods were compared: uncompressed, EightFold (design is divided into copies of eighths), METAC, and METAC16 (maximum number of levels is restricted to 16, for better compatibility with existing software). File size is plotted as a function of device diameter. (d) The scaling order, $b$, of the file size for each method is plotted. Error bars represents one standard deviation. Methods with $b$ closer to 2 indicate file size scaling with device area, while those with $b$ closer to 1 indicates scaling with device diameter.

Figure 2-10. Metasurface lens section designed using METAC, implemented using GDSII layout file format.

Figure 2-11. SEM of the center of a large metalens designed with METAC. The posts are made of amorphous silicon.

Figure 2-12. Metalens production by stepper. Metasurface lenses (metalenses) can be produced at low cost and high yield using existing photolithographic stepper technology. Here a wafer substrate is first deposited with the appropriate film stack, comprised of the metalens material (amorphous (a)-Si), photoresist (SPR700-1.0), and contrast enhancement material (CEM). The pattern of the metalens, which is contained in the reticle, is then projected by the stepper, and replicated rapidly over the face of the wafer by repeatedly exposing and incrementally stepping the wafer position. Throughputs as high as hundreds of wafers per hour (wph) can be
achieved. Then the pattern is etched into the a-Si, forming the metalens. Finally, after any residual photoresist is removed, the wafer can be diced into separate individual metalens devices. A photo of fabricated metalens (upper right), 2-cm in diameter, using this methodology is shown in comparison to a ruler. A SEM of the metalens center (center right) shows the microscopic posts comprising the metalens. Scale bar: 2 µm.

Figure 2-13. Flatness requirements. (a) As metalens size increases, the flatness over the device surface becomes increasingly relevant. The flatness requirements were studied by computer simulations of metalenses (diameter: 2 cm, focal length: 50 mm, design wavelength: 1550 nm), in which the surface curvature was varied (spatial period ranging from 1 to 100 mm, and perturbation amplitude from 0 to 200 µm). The intensity of the optical field is shown. The metalens is situated at the bottom of each plot, and the vertical (z) and horizontal (r) axes are the propagation direction and radial dimension, respectively. Each group four columns on the left and right show the optical behavior for even (cosine) and odd (sine) spatial perturbations, respectively, with respect to the metalens center. Quality of focus is reduced for shorter spatial periods and higher perturbation amplitudes. In (b), the surface profile of the 4-inch wafer we used (including metalenses) was measured (using Toho FLX-2320-S) and the Fourier transform calculated in (c) to obtain the major contributions to spatial frequencies, which mainly occurred at \( \Lambda < 0.02 \) mm\(^{-1}\) (or \( \Lambda > 20 \) mm). The inner and outer white circles denote \( \Lambda \) at 50 and 20 mm, respectively.

Figure 2-14. A schematic diagram showing the setup for characterization of the focal spot. A horizontal microscope and a camera are mounted on a motorized stage to scan the image the focal spot, which later can be 3D constructed to determine the focal length and the spot size of a metalens device.

Figure 2-15. Focus and imaging performance. (a) The flatness of the device allows for imaging setups very similarly to the ideal thin lens equation, which was used to demonstrate imaging capabilities. (b) Image of focal spot with 7 mm gaussian illumination with \( \lambda = 1550 \) nm. (c) The measured modulation transfer function (MTF) from (b) is plotted with the theoretical diffraction-limited MTF. Error bars: standard deviation. (d) Chromatic focal shift was measured as a function of the wavelength of illumination. The measured deviation of focal length from that of the design wavelength at 1550 nm (light blue dots, error bars: standard deviation) is plotted with the linear fit (blue line). (e) Hyperspectral image of focal spot in the same configuration as (b) for \( \lambda = 1440-1590 \) nm in 10 nm increments linearly binned to RGB channels (center wavelengths \( \lambda_R = 1590, \lambda_G = 1515, \) and \( \lambda_B = 1480 \) nm). The spot, which is largely white, indicates little chromatic aberration, which can be attributed to the low NA (0.07). Horizontal and vertical line cuts at the RGB center wavelengths are also shown. Using the thin lens setup in (a), simple imaging was demonstrated at \( \lambda = 1550 \) nm for (f) the Harvard university logo and (g) US Air Force 1951 resolution target, without any additional optical components. Scale bar (yellow): 1 mm.
Figure 2-16. Moore’s Law and metasurfaces. Being an early enabler of metasurfaces, Moore’s law, which predicts the transistor areal density in computer chips to double every year, is driven in large by improvements in lithographic technology. The plot shows the state of lithographic technology, represented as technology node size (TNS, diamond symbols), as a function of year, where smaller TNSs indicate higher feature densities. Several important product landmarks utilizing these TNSs are labelled, and key leaps in TNS powered by new light sources are denoted by vertical stems with circles. These past developments will enable the possibility of large area, cost-efficient metasurfaces optical devices. The red dotted line denotes the 700 nm threshold (corresponding to red light), below which the TNS had already surpassed in the mid-1990s. The wavelengths of the visible spectrum corresponding to TNS are shown shaded from 400-700 nm. In general, subwavelength-sized features can be produced using TNSs with at least twice the feature density as compared to the wavelength of light. Subsequent improvements to TNS up to the present (blue shaded region) have enabled feature sizes much smaller than the wavelength of light, providing ways to create more complex, fine-featured meta-elements as well as alternative routes for effectively utilizing foundry equipment which would otherwise be viewed as obsolete.

Figure 3-1. Principle of strain field-mediated tunable metalens. A metasurface (left column) is constructed by digitizing an analog optical phase profile on a flat surface into discrete cells, each of which contains a metasurface element that locally imparts the required phase shift to the incident light in order to reconstruct the desired wavefront (middle column, dotted line: optic axis). The wavefront generated by the metasurface determines the subsequent beam shaping (right column). (a) Original: metasurface without stretch. (b) Defocus: metasurface with uniform and isotropic stretch. (c) Astigmatism: metasurface under asymmetric stretch. (d) Shift: metasurface displaced laterally in the x,y plane.

Figure 3-2. Uniform stretch approximation. (a) The ideal stretch profile ($s_{ideal}$, which is the stretch profile that produces an aberration-free hyperboloidal phase profile) and the nominal, uniform stretch value ($s$) is plotted for typical values of $s$ from 1.01 to 1.3 (they are the same for $s = 1$) as a function of radial position. (b) The difference, $\Delta s = s_{ideal} - s$, is plotted for typical values of $s$ from 1 to 1.3 (blue to red lines, respectively), as a function of radial position for an example lens with focal length of 50 mm. (c) The magnitude of the wavefront aberration function (WAF) in radians (color bar), here defined as the difference between stretched ($\varphi_{stretched}$) and ideal ($\varphi_{ideal}$) phase profiles, is plotted as a function of radial position ($r$) and stretch ($s$) for a lens with focal length of 50 mm using light of wavelength 1550 nm. The white, dotted line shows the maximum aberration, which occurs at a stretch of $s \approx 1.22$.

Figure 3-3. Optical measurements of DEA, consisting of pre-stretched (4x) acrylic elastomer (VHB 4905, 3M) and single-walled carbon nanotube (SWCNT) electrodes. The measurements were performed using the Agilent Cary 7000 Universal Measurement Spectrometer and the Bruker Lumos FTIR microscope. The spectrum range is from 200 nm to 16 µm. The x-axis of the plots is in log scale. Our device operates at 1550 nm. Top: Transmission (blue), reflection (green), and
Figure 3-4. Dielectric elastomer actuator (DEA) metalens device design. (a) A schematic of the device in which a metalens and DEA with 5 addressable electrodes are combined to allow for electrical control over the strain field of the metasurface. (b) Optical microscope images (scale bar: 20 µm) at (i) no voltage, (ii) 2.5 kV applied to the center electrode (V5), and (iii) 2.75 kV applied to tune X-astigmatism (concurrently, V1 and V3). The dark spots are defects (either missing or tilted silicon posts) introduced during the transfer process. The corresponding 2D Fourier transforms of the images (i)-(iii) are shown in (iv)-(vi), respectively (normalized amplitudes).

Figure 3-5. Metalens design elements. (a) The unit cell of the lens consisting of the SiO2 substrate and an amorphous silicon (a-Si) post. The dimension are labeled as follows: h is the height of the post, d is the diameter of the post, u is the size of the unit cell and e is the constant edge-to-edge spacing between neighboring posts. (b) Phase and amplitude response of posts are plotted as a function post diameter. (c) The metasurface design of our device is shown. Left: The whole lens is 2 cm in diameter and appears as a solid black circle in this rendering due to the sheer density of structures. Top right: a close up view showing a 200 µm wide window of the center of the lens. Bottom right: a close up view showing a 50 µm wide window of the center of the lens.

Figure 3-6. Polyacrylic acid (PAA) based sacrificial layer. PAA was attempted to be used as the water-soluble sacrificial layer. Metalenses are shown with damaged film on a hot plate. High temperature processing steps as well as exposure to humidity and water caused difficulty in the process (in particular, from thermal expansion of the polymer).

Figure 3-7. Dielectric elastomer actuator platform. (a) Fabrication steps: (i) Schematic of the thin-film stack prepared for photolithography. (ii) After patterning and etching, the stack is attached to an elastomer layer and the whole stack is immersed in a water bath. (iii) Schematic of the release process showing the dissolution of the sacrificial layer (GeO2) from the outer edge of the device towards the center, leaving the metasurface structure attached to, and supported by, the elastomer layer. (iv) A schematic of metasurface and DEA combination in which an applied voltage supplies the electrode layers (SWCNTs) with electrical charges. Their electrostatic attraction compresses the elastomer in the thickness direction, and causes expansion in the lateral direction. (b) False-colored scanning electron microscope image of the center of the lens shows a-Si posts on the GeO2 sacrificial layer before attaching to the elastomer layer.

Figure 3-8. Dissolution of a GeO2 based sacrificial layer in water, releasing a thin film of silicon.
Figure 3-9. Device design and operation. (a) The device with the metasurface at the center of a pre-stretched DEA sheet, suspended by a plastic frame. The electrical contacts are made to the SWCNT electrodes using silver grease and carbon tape (black stripes). There are eight connections using SWCNT, labeled V₁-V₅ and GND (ground). Schematics (cross section) of single layer (SL) and double layer (DL) devices are shown. Both devices have the same basic structure in which the metasurface is attached to a DEA via an intermediate elastomer layer (IEL). (b) Different electrode activation configurations. Applying a voltage to an electrode will cause the area under that electrode to expand. The defocus (+) tuning is controlled by the center electrode, while (-) tuning is controlled by applying voltages of equal magnitude to four peripheral electrodes. Astigmatism (X or Y) is controlled by opposing pairs of peripheral electrodes. The X, Y shift is controlled by activating one of the peripheral electrodes. .......................................................... 61

Figure 3-10. Metalens and dielectric elastomer actuator (DEA) mating. First fabricated separately, the (a) metalens that is transferred to an elastomer substrate is combined with (b) the DEA to form the (c) adaptive lens device. ......................................................... 63

Figure 3-11. Transfer process. (a) A round silicone press was used for making initial contact between the metalens on wafer to the elastomer substrate. (b) An apparatus held the silicone press and a controlled force was applied for several minutes to hours. (c) Hand pressing was also used in some cases. (d) The lens and elastomer combination was then visually inspected for any air bubbles or other defects. (e) The combination was then immersed in water so that the sacrificial layer could be released. (f) With the wafer substrate removed, the lens was then completely transferred onto the elastomer substrate. ......................................................... 64

Figure 3-12. SEM of released posts. The sample was briefly exposed to water and then dried to visualize the released (as well as unreleased) posts. ................................................. 65

Figure 3-13. Measurement setup. Schematics (top view) of the setups for measuring the focused spot size of the device (a) with and (b) without the horizontal microscope to magnify the image. The setup without the horizontal microscope was used in at higher applied voltages, in which there was not enough space on the motorized stage to accommodate the horizontal microscope given the extended focal length of the device. (c) Setup for measuring the focusing efficiency of the device. The first aperture was used to reduce the incident beam size to 6 mm to match with the size of the device. The second aperture was used to exclude the stray light. The focusing efficiency was calculated using the ratio of the optical power at the focus (measured with both apertures and the device in place) to the total unobstructed optical power (measured with both the device and second aperture removed, but with the first aperture still in place). ......................................................... 67

Figure 3-14. Measurement of tuning. (a) (Left) Z-scan of the intensity profile showing two distinct focal lengths. (Center) Image of focal spot intensity profile, i.e. x-y cross sections at the position of maximum intensity. (Right) Line scans of focal spot intensity image in the x (blue) and y (red) directions in comparison to the theoretical diffraction-limited spot size (black). (b) Additional images of the focal spots at 1.5,
2, and 2.5 kV are shown, captured directly by the camera without the microscope objective (scale bar: 200 µm). (Lower right) Measured focusing efficiency for varying voltages. (c) Measurement of focal length tuning using center electrode V5 for double layer (DL) and single layer (SL, inset) devices. Blue circles: Optical measurement of device focal length as a function of applied voltage. Solid blue line: fit of focal length data to Eq. 3.3 (R2 = 0.9915). Red triangles: Measurement of stretch as a function of the applied voltage. (d) Measured Zernike coefficients of the phase profile (calculated from microscope images of lens center) showing targeted tuning of vertical astigmatism, while other Zernike coefficients exhibit little change. The large defocus value represents the designed focal power of the lens. (e) Measurement of x,y-shift control, showing two-axis displacement control from the focus position at 0 kV (yellow star) to various positions (red dots) as 1.9 kV is applied. The gray shaded area shows possible displacements, which can be accessed by concurrently activating combinations of electrodes.

Figure 3-15. Fitting measurement of focal length tuning using electrode V5. Fitting measured focal length with applied voltage for (a) SL device and (b) DL device. Black dots: Measurement of device focal length as a function of applied voltage. Blue solid lines: fits of the relationship between the focal length and the applied voltage. Blue dashed lines: fits with 95% confidence bounds.

Figure 3-16. XY-shift distortion measurement. XY-shift distortion was measured by using microscope images of metasurfaces, for which displacements were produced in the up, down, left, right ((i)-(iv), respectively, with scale bar: 20 µm) directions by applying 1.9 kV to the four peripheral electrode areas. Shown below ((v)-(viii)) are the corresponding Fourier transforms of (i)-(iv), respectively. During imaging, the microscope field of view was re-adjusted with respect to the center of the metasurface after each voltage application.

Figure 3-17. Response time measurement. Measured rise and fall times for DL (a-d) and SL (e–f) devices. The measured data are plotted as black dots, the fits to exponential curves are shown as blue solid lines, and 95% confidence interval of the fit are shown as blue dashed lines. DL was actuated with a step voltage between 0 and 2 kV showing a (a) rise time of 33±3 ms and (b) fall time of 271±3 ms. DL was actuated with a step voltage between 0 and 3 kV showing a (c) rise time of 182±15 ms and (d) fall time of 105±13 ms. SL device actuated with a step voltage between 0 and 2 kV showing a (e) rise time of 327±7 ms and (f) fall time of 320±8 ms.

Figure 4-1. Three Stokes parameters form the axes of the Poincaré sphere.

Figure 4-2. Concept of parallel polarization state generation. (a) An illustration showing the general, modular implementation of the described method for a parallel polarization state generator (PSG). An input beam is (i) split into four beams of different polarizations, which are then (ii) intensity modulated either in reflection or transmission, (iii) and finally combined to form a single output beam, the polarization and phase of which can be tuned with a precision and speed limited by the modulator. (b) A schematic of PSG architecture is shown, in which modulators are placed after light sources Ai with well-defined states of polarization (SOP) and
relative phase, and their weighted linear superposition produces the desired output signal. (c) Generation of horizontally polarized light using this method is illustrated. The electric fields of four propagating electromagnetic waves (red, green, blue, and yellow) with elliptical polarizations are superimposed and plotted as function of wave propagation position. They are intensity modulated and beam combined to generate the desired horizontal polarization (black).

Figure 4-3. Simulations of polarization state coverage. Two systems with distinct sets of Stokes basis vectors (SBVs) were simulated— one composed of degenerate SOPs and the other with SOPs mapped to a regular tetrahedron on the Poincaré sphere. Intensity modulation parameters of individual SBVs are varied to generate polarization trajectories and whole regions of accessible SOPs. (a) A system with SBVs with four degenerate SOPs: linear horizontal (C1), vertical (C2), +45° with a 180° phase shift (C3), and right circular polarization (C4), are shown. A Monte Carlo simulation (blue points) was performed by randomly varying the intensity modulation parameters and showed complete, yet non-uniform coverage of SOPs over the Poincaré sphere. A polarization trajectory between SBVs C3 to C4 is shown for coherent combination (blue line) and incoherent combination (red line). Incoherent trajectories are geodesics. (b) A system with SBVs optimized for better uniformity of SOP coverage is shown, in which the SBVs are vertices of a regular tetrahedron inscribed in the Poincaré sphere, as opposed to the degenerate SBVs. In Jones vector notation, the SBVs used here were [0.7071, 0.7071i], [-0.856, 0.1691i], [0.5141, 0.7941-0.3242i], and [0.5141, -0.7941-0.3242i], labeled as C1-4, respectively.

Figure 4-4. Phase dependence of coherent polarization trajectories on the Poincaré sphere. The effect on polarization trajectories by changing the relative phase between two linearly combined SOPs is shown. The polarization trajectory connecting SOPs can be modified in either direction by controlling the relative phase.

Figure 4-5. Regions of coverage spanned by Stokes Basis Vector subsets. (c) The degenerate SBV system is mapped using a Mercator projection of the Poincaré sphere, where θ is the polar angle and ϕ is the azimuthal angle. All coherent and incoherent polarization trajectories between SBVs are shown with black dotted and red solid lines, respectively. The coherent polarization trajectories connected to C1 are warped by increasing the relative phase difference between C1 and other SBVs by 6° (blue dotted lines). The colored regions show the regions of SOPs enabled by combinations of three SBVs: by combining C1, C2, and C4, with varying intensity modulation parameters, all SOPs in the blue region can be generated; similarly, combinations of (C1, C3, C4) and (C2, C3, C4) generate the red and green regions, respectively. However, (C1, C2, C3) generate a region of no area because these SBVs are not linearly independent in this particular system. (b) The regular tetrahedron SBV system. Coherent and incoherent polarization trajectories between SBVs are shown with black and red dotted lines, respectively. In this case, regions of SOPs generated by combinations of sets of three SBVs are well distributed and have similar size and great overlap, yielding better uniformity. The overlapping
regions are color coded and labeled as the following: C1, C2, C3 combine to cover regions a, b, and c; similarly: C1, C2, C4 (a, d, e); C1, C3, C4 (c, e, f); and C2, C3, C4 (b, d, f).

Figure 4-6. Degenerate SBV coverage. The Poincaré sphere is shown, covered by possible SOPs, as generated by linear combinations of four degenerate SBVs, in the following polarizations: linear horizontal, linear vertical, linear +45°, and right circular. All SBVs have global phases $\phi=0^\circ$, except that of the linear +45° polarization with $\phi=180^\circ$.

Figure 4-7. Coherent polarization trajectories on the Poincaré sphere. Trajectories are shown, where the polarization is varied from the initial state (red) to the final state (blue). Four trajectories between degenerate SOPs are plotted: a) linear horizontal to linear vertical, b) right circular to linear horizontal, c) linear vertical to linear +45°, and d) linear +45° to right circular. The global phase of the linear +45° polarization is defined to be 180° out-of-phase with respect to the other degenerate SOPs, such that in Jones vector notation it is [-1, -1].

Figure 4-8. Coherent and incoherent polarization trajectories on the Poincaré sphere. a) Various perspective views of an example system. The SBVs used are labeled C1-4. Polarization trajectories are generated by coherent combination (blue line) and incoherent combination. b) This example system is implemented using the experimental setup described in the main text. A polarization trajectory is generated by coherent combination by keeping the optical path length between the two SBVs C2 and C3 well below the coherence length of the laser (~20 cm). The trajectory is measured by the polarimeter and the data are shown. c) An incoherent polarization trajectory following the geodesic path between SBVs C2 and C4 is generated by making the optical path length between SBVs much longer than the coherence length of the laser (hence reducing the mutual coherence).

Figure 4-9. Experimental setup. Laser light is prepared in the linear +45° polarization using a wire-grid polarizer. The beam is split into two beams by a non-polarizing beam splitter (BS). Each of these beams is split again using variable circular polarizers (VCPs) into elliptical polarization states, which can be tuned by rotating the internal quarter wave plates. Variable neutral-density filters (VNDFs) are placed directly after the VCPs to balance the four beam intensities. The four beams are then directed onto four quadrants of the surface of a computer controlled Texas Instruments DLP3000 digital micromirror device (DMD). The DMD is composed of an array of polarization-insensitive mirrors that can be switched in one of two positions. Mirrors that point in the direction of the output beam contribute to the total intensity and all other light is directed into a beam dump. The DMD behaves as a 2-D diffraction grating for the incident laser light. An iris is used to select the strongest diffraction order. The path length differences of the four intensity-modulated beams passing through the iris are adjusted to be less than the coherence length of the laser (< 20 cm) with a series of mirrors. They are combined using three non-polarizing beam splitters to form a single beam. Finally, this beam is passed through a 100-µm pinhole, in order to select a small uniform portion of the beam.
wavefront of the combined beam to maximize the degree of polarization, to form the PSG output.

Figure 4-10. Experimental results. (a) The Stokes basis vectors (SBVs) are set to SOPs approximating a regular tetrahedron on the Poincaré sphere. The SBVs C1-4 were measured and the resulting tetrahedron is drawn. Coherent polarization trajectories between all SBVs are generated by modulating intensities in 20 discrete increments spanning 20 seconds, and the data as measured by polarimeter are shown. (b) The results of a Monte Carlo experiment, in which 200 random intensity modulation parameters $\alpha$ were used, are shown on the Poincaré sphere, indicating good uniformity of coverage of SOPs. (c) Time series data of a coherent polarization trajectory between two SBVs (C2 to C4) in (a) are compared to theoretical calculation (dotted line) and show good agreement. S1, S2, and S3 are elements of the Stokes vector. (d) An eye pattern is generated for a polarization signal that switches between linear horizontal and vertical polarizations using the DLP3000. The data are shown for a pseudorandom bitstream modulated at 1 kHz. The inset shows the measured settling time (eye rise and fall time) to be 3.5 $\mu$s following an exponential.

Figure 4-11. Insertion loss calculation. Insertion loss is calculated for 80,000 SOPs distributed uniformly over the Poincaré sphere. The distribution of insertion losses are represented by histograms, for which a set of 4 degenerate SBVs and a set of regular tetrahedral SBVs are shown in blue and orange, respectively.
Acknowledgments

I thank my teachers, colleagues, friends, family for their support.

First, I thank my advisor, Professor Federico Capasso, who has been a great mentor and friend. I really appreciate your teaching method, scientific insight, and the group research environment you have created. Thank you also for accompanying me through the ups and downs of my PhD. It is hard to describe what is the most valuable thing I have learned from you, because there are so many… here a few that I immediately think of: in terms of leadership – to lead by example; in terms of research - to pierce through the murkiness.

Secondly, I thank Professor David Clarke, who first accepted me into Harvard and has also been a great mentor and friend. I appreciate how you have always kept an interest in my work, even after I moved to Federico’s group, and provided valuable insights along the way. It was a great pleasure to have been able to work together on my last project.

In my dissertation committee, in addition to Federico and David, I would also like to thank Professor Marko Loncar, for valuable advice and suggestions. I also thank
Professor Evelyn Hu, who was on my qualifying committee, for valuable advice and comments.

I thank Sam Shian for your friendship and a wonderful collaboration on the adaptive metalens project. It would not have succeeded without you nor would it have been so enjoyable.

I thank Roger Diebold for your friendship and scientific advice, which you were willing to provide even at your busiest moments.

I would also like to thank my collaborators in other projects that I was apart of that were not included in this thesis: Vincent Ginis, Lulu Liu, Patrice Genevet, Daniel Wintz, Antonio Ambrosio, Tobias Mansuripur, and Romain Blanchard. I thank the rest of the Capasso group and all its members for a great working environment.

I would also like to thank my co-advisor, Y.K. Chen, and colleagues at Bell Labs during the SPA project, who provided a extremely enriching and fun place to work. I thank Y.K. for your kind guidance, friendship, and scientific insights. I thank Guillem de Valicourt for your friendship and making research an excitement-filled drag race, which caused me to lose quite a lot of sleep. I thank Chia-ming Chang for your friendship and always coming over from the office next to mine to discuss our silly ideas. I thank Po Dong for always making me think.

I thank CNS and its staff at Harvard for the facilities to “make the little stuff.” I especially thank Guixiong Zhong, who is not only a close friend but also helped me greatly with nanofabrication. I thank Yuan Lu and JD Deng for your continued support through these years. I also thank Ling Xie, Ed Macomber, and Jason Tresback with your help in cleanroom fabrication.
I thank CNF and its staff at Cornell. I especially thank John Treichler, who tirelessly helped me overcome challenge after challenge in making high data density photomasks. I also thank Gary Bordonaro, Aaron Windsor, Karl Musa, David Botsch, Chris Alpha, and Mike Skvarla for their fabrication guidance and help in settling into the CNF cleanroom.

I thank Draper for financially supporting my PhD. I thank JP Laine for being my Draper Fellowship adviser and making it easy. I also thank Steve for his guidance.

I thank Dudley House and the people who run it, especially Susan, Jeff, and Jim for a place to go (outside of lab). I thank the Dudley Fellows for the good times.

I thank my teachers from before, without whom I would not be doing what I am doing today: Dr. Craxton from LLE and Professors Gianluca Gregori, Todd Huffman, and David Andrews from Oxford, Professor Kazuo Tanaka from Osaka U, and Jason Pronyk from the UN.

I thank my friends at Harvard: Anna, Bobby, Cheng, Chuanxi, Danqing, Fei, Hanrong, Jing, Jingyee, Longzhi, Nanxi, Ray, Ruobing, Tina, Wenbo, Xizhu, Yang, You, Young-ik. I thank my friends elsewhere who supported me during my PhD: thank you Dennis, Seward, Jack, Kelvin, Harrison, Austin.

I thank my mother, father, and sister for their unconditional support. I also thank my extended family for their support.

Last but not least, I thank my wife, Shuyan Zhang, for your love and support, and allowing me to have the great pleasure and unexpected opportunity of concocting ideas and doing experiments together with the person I love. Getting through the PhD, I owe to you.
Dedicated to

my wife, parents, and sister
Chapter 1
Introduction

1.1 Properties of light

Throughout human history and to this day, light has been a subject of mystery and intense study. From more than two millennia ago, what began as philosophical musings concerning the nature of light has been debated, explored, and more recently tested by experiment through the scientific method. From the particle (“corpuscular”) theories of light suggested by Pierre Gassendi and Isaac Newton in the 16th and 17th centuries, to the wave theories proposed by Robert Hooke, Christiaan Huygens, and Leonhard Euler in the 16th to 19th centuries, then to the electromagnetic theory introduced by Michael Faraday, James Clerk Maxwell, and Heinrich Hertz in the 18th century, and finally to the quantum theory of light presented by Max Planck and Albert Einstein in the 20th century (to which we adhere to this day), our concept of light has undergone many changes (Hecht 2002).

Today, we understand light to be an oscillation of the electromagnetic field, which may carry radiant energy. In the current theory, based on quantum mechanics, light is more mysterious than ever before – simultaneously expressing the behaviors of both waves and particles. Furthermore, the speed of light is taken to be one of the fundamental constants of nature, because it is independent of the inertial motions of both sources and
observers, thus forming one of the basic assumptions of special relativity. In the face of these curious notions, we may be able to claim, perhaps, that over the years we have slowly unriddled some of light’s many secrets and gradually acquired a means of control over some of its aspects.

Light ordinarily refers to the visible portion of the electromagnetic spectrum that provides the human eye with the sense of vision. In physics, however, it can refer to electromagnetic radiation in any part of the spectrum. The visible spectrum (wavelengths between 400 to 700 nm) is neighbored by the infrared and ultraviolet spectra, which correspond to wavelengths that are longer and shorter, respectively, than the eye can see. However, the fundamental properties of light are the same throughout the electromagnetic spectrum and can be defined with the parameters of intensity, wavelength, shape of the wavefront (which dictates the propagation direction), and polarization. All of these are important parameters that dictate the behavior of light and form the basis for many key technologies used in everyday life. This thesis focuses on the latter two: wavefront and polarization. The ability to precisely control the wavefront and polarization as a function of time is highly desirable, as it is essential for numerous powerful modern applications, including imaging, high tech manufacturing, and communications, to name a few. There has been a bulk of work on the manipulation of these properties, but there remains much to be improved in terms of extreme compactification, reduction of power, and increase in speed and precision, as everyday use cases are driven towards untethered, lightweight, and high performance devices. In this thesis, novel devices are presented that dynamically shape wavefront and polarization of light, using new principles of operation and device designs.
Figure 1-1. Wave nature of light. Light is shown as a transverse electromagnetic wave with wavelength of $\lambda$ and propagating in space ($x$) in the direction of its wavevector, $k$. The electric and magnetic field components are shown in blue and red, respectively.

The electromagnetic nature of light is usually depicted as in Figure 1-1. When travelling in free space, light occurs as a transverse electromagnetic wave (i.e. the electric and magnetic field directions ($\mathbf{E}$ and $\mathbf{B}$, respectively) are perpendicular to each other) and propagates in time in the direction of its wavevector, $k$, which is perpendicular to both $\mathbf{E}$ and $\mathbf{B}$ and has a magnitude given by $k=2\pi/\lambda$. Here, the polarization of light refers to the direction of the electric field oscillation. While useful, this depiction does not account for the fact that light is a field that extends through three-dimensional space and propagates in time, which is also responsible for diffractive phenomena. Figure 1-2 better conveys spatially extended nature of light. The locus of points in space that share the same phase (“iso-phase”) draws out the wavefront of light, much like the profile of the crests or troughs of an ocean wave. The shape of the wavefront determines all of the local directions in which the wave travels as well as the direction of propagation of radiant...
Figure 1-2 Illustration of the wavefront and polarization of an example light wave extended in two dimensions, x and y, and polarized (red arrows) in the out of plane direction, z. The locus of points with the same phase (“iso-phase”) illustrates the wavefronts of light – here they are shown as black dotted lines. The direction perpendicular to the wavefront corresponds to the local propagation direction of the light wave (k).

energy. For example, a wavefront that matches the shape of a circular arc may cause optical energy to travel towards a single point in center of the circular arc, giving rise to the most common application of wavefront shaping: lenses.

1.2 Shaping the wavefront of light

Lenses are optical devices that use curved surfaces, traditionally made of glass, to focus light by shaping its wavefront (Hecht 2002). From a geometric optics point of view, the refractive index of glass, which differs from air, together with the curved surface of
Figure 1-3. Lenses as wavefront shapers. Lenses and many other optical devices, in general, can be considered as objects whose primary functions are to transform the incident wavefront into a desired outgoing wavefront. Here a thin lens, represented as a thin yellow line, focuses an incoming plane wave ($\varphi_{\text{in}}=2\pi x/\lambda$) to a focal point (where $x_0$ is the focal length) by transforming the planar wavefront into a spherical one ($\varphi_{\text{out}}=2\pi[x^2+y^2]^{1/2}/\lambda$). The lens function ($\varphi_{\text{_lens}}$) is then simply the phase profile that provides the phase shifts that brings $\varphi_{\text{in}}$ in phase with $\varphi_{\text{out}}$: $\varphi_{\text{in}} + \varphi_{\text{_lens}} = \varphi_{\text{out}} + \text{constant}$. Since $\varphi_{\text{in}}$ provides only a constant phase offset in the $y$-direction of the lens, it can be neglected, thus giving: $\varphi_{\text{_lens}} = \varphi_{\text{out}}(x_0)=2\pi[x_0^2+y^2]^{1/2}/\lambda$.

The lens act to bend light rays towards or away from a focal point, in which the bending angle is given by Snell’s law. This can be equally well explained using wave optics: the profile of the lens generates a profile of accumulated optical phase as light travels through it. Therefore the lens imposes a spatial profile of phase shifts (i.e., “phase profile”) given by the thickness of the glass, such that the wavefront is then correspondingly curved as the light exits the lens, which bends the light towards or away from the focal point. Figure 1-3 illustrates the concept, in which a lens is an object that
Figure 1-4. The human eye. The human eye is an advanced imaging system that is capable of adapting the shape of lens using small muscles to focus light from objects at different distances onto the retina. Image credit: By Jmarchn (Own work) [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons.

transforms light from an incoming wavefront to an outgoing wavefront: in this case, a plane wave to a spherical wave, respectively, through a specifically designed phase profile.

In fact, the human eye (Atchison & Smith 2000; Oyster 1999) is an excellent example of an advanced lens. The eye is a dynamic wavefront shaping system, which contains a deformable lens. By adjusting the shape of the lens using small muscles that are connected to the lens through ligaments, the wavefront of the light that passes through it is also correspondingly shaped. This operational principle inspires the adaptive lens
presented later in this thesis. In general, changing the shape of an optical intermediary will change the shape of the interacting wavefront. Importantly, it is the shape in “optical phase space” (i.e., geometric dimensions given by optical path lengths, which account for refractive indices) that ultimately matters – in many other implementations of wavefront shaping devices, in particular, man-made devices, such as liquid crystal spatial light modulators (Itoh et al. 2009), the shape of the device appears to be unchanged far as we can see, but at the microscopic level, there are small changes that result in a change in shape in optical phase space.

1.3 Metasurfaces

Wavefront shaping is performed in this thesis using metasurfaces. Recently, there has been significant scientific and technological interest due to their inherent multifunctionality and potential to dramatically reduce the thickness of the optics in a wide range of applications (Yu & Capasso 2014). In essence, metasurfaces are similar to phased arrays. Phased array technology (Visser 2005) was traditionally used in the radio spectrum and was invented in 1905 by Nobel laureate Karl Ferdinand Braun (Nobelprize.org 2017; Braun 1909) to demonstrate enhanced transmission in one direction. The control of the direction of transmission is achieved with a large number of antennas, usually spaced very closely and situated on a common planar surface. By controlling the phase shift of each individual antenna, a phase profile over the surface can be imposed, and thus the outgoing wavefront can be defined. Phased arrays are important in many areas, including radar, radio broadcast, and communications, for which the transmission direction can be controlled or rapidly scanned and shaped by electronically
controlling the phase profile. Metasurfaces are phased arrays that (a) have been adapted to shorter wavelengths, most commonly in the visible and near infrared wavelengths, and (b) are not powered nor do they actively emit radiation in current implementations but rather only interact passively with incident radiation.

In order to have phased array with high transmission efficiency, a high density of antennas is desired (Mailloux 2005), i.e. the spacing between antennas should be less than the wavelength (“subwavelength”). In many implementations, antennas are spaced by a quarter of a wavelength. For small antenna spacing, the phased array emulates an object with a continuous phase profile; whereas for large antenna spacing, the discretization of the phase profile is more pronounced, such that the phased array assumes the characteristics of diffraction gratings – the appearance of grating lobes, into which energy is distributed. Usually, grating lobes other than the main lobe are undesirable, and the energy that is directed into them is treated as a loss in efficiency.

At radio frequencies, it is relatively easy to make subwavelength-spaced antennas, since the wavelengths in question range from as long as hundreds of kilometers to as short as millimeters. At the shorter wavelengths that concern metasurfaces, such as the near infrared and visible wavelengths, the antennas must be spaced by at most a single micron or only several hundred nanometers, respectively, in order to satisfy the subwavelength criterion. Due to the difficulty in creating such small antennas, phased arrays have not been scaled down to these sizes until recently.

The ability to make antennas at this scale has been brought about by advances in the semiconductor industry, in which the techniques of nanofabrication has been applied to making optical devices, spurring much activity in a relatively new field:
nanophotonics. Here, the same top-down fabrication processes used in making integrated circuit chips, such as photolithography and electron beam lithography, are used to make nanoscale structures for the purposes of manipulating the material interaction with light. Lithographic processes can be used to make patterns of whole arrays subwavelength-spaced antennas, which introduce a desired spatial distribution of optical phase, amplitude, and/or polarization. By tailoring the properties of each element of the array, one can spatially control these properties of the transmitted, reflected, or scattered light and consequently mold the wavefront. Generally, one can imagine arbitrary phase profiles to generate a wide variety of optical functions. Based on this concept, various functionalities have been demonstrated including lenses, axicons, blazed gratings, vortex plates and wave plates.

One of the remaining challenges, however, in making practical devices with metasurfaces is fabricating them with large areas. The key obstacles in this endeavor is the enormous data density required by the subwavelength resolution criterion, which must be maintained over large areas. This results in unmanageably large file sizes for the layout design files describing structures greater than a few hundred microns in diameter, which can often be greater than 100 gigabytes. In an effort to solve this issue, Chapter 2 presents a scalable metasurface layout compression algorithm that exponentially reduces the design file size (by 3 orders of magnitude for a centimeter diameter metalens, and even greater gains in compression for larger sizes) and a route to mass manufacturing of metasurface lenses (metalenses) using stepper photolithography with extremely large areas, up to (but not limited to) centimeters in diameter (She et al. 2017b). Because of what is demonstrated here, the claim that chipmakers will be making lenses of the future
(as well as the chips), i.e. the unification of two industries: semiconductor manufacturing and lens-making, is now closer to reality.

However, metasurfaces are by themselves static, such that additional steps are required in order to introduce the feature of tunability, e.g. focal length or magnification control. To take advantage of the thinness and planarity of metalenses, electrical tuning of lateral motion is essential for focal length and magnification control, which is performed by longitudinal mechanical motion along the optical axis in conventional optical systems. Using the large area metasurface technology developed in Chapter 2, a large area, integrated optical device, no more than 30 microns thick is presented in Chapter 3. This device imprints a strain field onto the optical wavefront, by way of a soft metasurface intermediary, enabling simultaneous control over focal length, astigmatism, and image shift (She et al. 2017a). These combined capabilities have so far been possible only in electron optics, and has been made possible by combining metalenses, artificial muscles (i.e., dielectric elastomer actuators (DEAs)), and carbon nanotube-based stretchable transparent electrodes. This is the first step of many that seeks to exert arbitrary control over the strain profile of a metasurface: transforming a transformation optic.

### 1.4 Shaping the polarization of light

The polarization of light is a property that defines the direction of the electric field component of the electromagnetic wave. Polarizations often are met in the linear or circular state – the electric field points to a single direction or rotate as the wave propagates, respectively. The state of polarization (SOP) can also be in any intermediate
state between linear and circular: elliptical (Collett 2005). Today, polarization is used in precise measurement techniques, such as ellipsometry of thin films, visualizing material stress, and imaging in low contrast conditions. However, the way that polarization most affects our everyday lives, is perhaps through optical fiber communications, which provides the world with the information highways connecting the Internet. Within optical fibers, as many as hundreds of different polarization states are being generated and switched between in rapid succession (through a combination of quadrature amplitude modulation (QAM) and polarization division multiplexing (PDM)), in order to encode and transmit information, obtaining speeds of gigabits to even terabits per second.

Polarization shaping is achieved using a new architecture, inspired by the mathematical form of the Stokes Parameters, which describe the SOP of light, including its degree of polarization (DOP). Previously, the general problem of generating arbitrary time-varying SOPs has always been mathematically formulated by a series of linear transformations, i.e. a product of matrices, imposing a serial architecture, such as a series of rotating wave plates or applying strain to a fiber. An alternative parallel architecture is presented, which is described by a sum, rather than a product, of matrices (She & Capasso 2016). The theory is experimentally demonstrated by modulating spatially-separated polarization components of a laser using a digital micromirror device (DMD) that are subsequently beam combined. This method greatly expands the parameter space for engineering devices that control polarization. Performance characteristics, such as speed, stability, and spectral range, are entirely dictated by the technologies of optical intensity modulation – absorption, reflection, emission, and scattering, any of which may be used. This opens up important prospects for polarization state generation (PSG) with
unique performance characteristics with applications in spectroscopic ellipsometry, spectropolarimetry, communications, imaging, and security.

1.5 Summary

The methods explored here have the potential of opening up a wealth of future new applications, including electrical control over focal length and polarization at high speeds in compact form factors. Furthermore, advanced tunability can be implemented, such as correcting for optical aberrations, which could lead to major advances in optical microscopes and imaging systems as well as portable and wearable devices. The results demonstrate the possibility of future optical microscopes, which operate fully electronically, as well as compact optical systems that rapidly probe various polarization states or employ the principles of adaptive optics to correct many orders of aberrations, simultaneously.
Chapter 2
Scalable enlargement of metasurfaces for practical purposes

2.1 Introduction

Nanofabrication, or the fabrication of devices with nanometer-scale features, is a group of manufacturing technologies brought about by the requirements of the semiconductor industry, in order to increase the number of transistors within a computer chip, keeping pace with Moore’s Law (Moore 1965; Mack 2015). Beyond its original purpose in electronics, it has caught the attention of a range of other industries and scientific fields, including optics and photonics, medicine, aerospace, among many others.

Optical components, such as lenses, have traditionally been made in the bulk form by shaping glass or other transparent materials to form smooth surfaces with which to bend light toward desired directions. Recent advances in metasurfaces provide a new basis for recasting optical components into thin, planar elements using the techniques of nanofabrication, having similar or better performance by using arrays of subwavelength-spaced optical phase-shifters (Yu et al. 2011). The technology required to mass produce them dates back to the early 1990s, when the feature sizes of semiconductor
manufacturing became denser than the wavelength of light, advancing in stride with Moore’s law. This provides the possibility of unifying two industries: semiconductor manufacturing and lens-making, whereby the same technology used to make computer chips is used to make optical components, such as lenses, based on metasurfaces (She et al. 2017b). Using a scalable metasurface layout compression algorithm that exponentially reduces design file sizes (by 3 orders of magnitude for a centimeter diameter lens) and stepper photolithography, we show the design and fabrication of metasurface lenses (metalenses) with extremely large areas, up to centimeters in diameter and beyond. Using a single two-centimeter diameter near-infrared metalens less than a micron thick fabricated in this way, we experimentally implement the ideal thin lens equation, while demonstrating high-quality imaging and diffraction-limited focusing.

2.2 METAC algorithm

To demonstrate large area metasurface optics using a process that could scale to high volumes and low costs, we designed and fabricated a transmissive metasurface lens (metalens) 2 cm in diameter. In order to produce a lensing effect, the metalens imposed a spatial profile of phase shifts (phase profile) (Wang et al. 2017) on the wavefront is \( \varphi = -\frac{2\pi}{\lambda}[(r^2 + f^2)^{1/2} - f] \), where \( \lambda = 1550 \text{ nm} \) is the design wavelength (1550 nm), \( r \) is the radial
Figure 2-1. Metasurface lens design. (a) Schematic of a metasurface lens (metalens) that is designed to focus light of normal incidence, where D is diameter and f is focal length. The phase profile (blue curve) is implemented by a dense array of microscopic meta-elements, made of amorphous silicon cylindrical posts, supported by a SiO$_2$ substrate. (b) The diameter of each meta-element is used to control its phase response. The phase response (blue) and transmittance (red) is plotted as a function of post diameter.

coordinate, and f=50 mm is the focal length (Figure 2-1a). The wavelength was chosen to according to the requirements of the equipment used in fabrication. Amorphous silicon (a-Si) on silica (SiO$_2$) was chosen as the metalens material. This phase profile focuses collimated light into a spot and was constructed using a dense pattern of meta-elements, each of which acted as a miniature antenna to locally impart a controlled phase shift. We used cylindrical posts, given their polarization-independent response, with diameters ranging from 830 to 990 nm and a fixed height of 600 nm (Figure 2-1b). By varying the diameter, the posts were able to produce $2\pi$ phase coverage and high, relatively uniform transmittances. To maximize optical efficiency, the placement of meta-elements was made denser using fixed edge-to-edge separation instead of the traditional fixed center-
Figure 2-2. FDTD simulations of the focal spot of a 2 cm diameter metalens: (a) Simulated distribution of the electric field intensity (normalized $|E|^2$) of the focal spot. (b) The cross section of intensity profile at $Y = 0$ (white dashed line in (a)) from which the size of the focused beam can be determined. This simulation indicates that the focal spot size, i.e. the full beam waist at $1/e^2$ of the peak intensity, is 17.1 µm.

<table>
<thead>
<tr>
<th>Device diameter</th>
<th>Meta-element count</th>
<th>Uncompressed file size (B = Byte)</th>
<th>METAC file size</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 µm</td>
<td>150</td>
<td>19.8 kB</td>
<td>9.1 kB</td>
</tr>
<tr>
<td>50 µm</td>
<td>3,614</td>
<td>157.9 kB</td>
<td>44.5 kB</td>
</tr>
<tr>
<td>100 µm</td>
<td>14,068</td>
<td>501.4 kB</td>
<td>94.1 kB</td>
</tr>
<tr>
<td>500 µm</td>
<td>1,053,822</td>
<td>30.3 MB</td>
<td>972.1 kB</td>
</tr>
<tr>
<td>1 mm</td>
<td>3,204,089</td>
<td>91.2 MB</td>
<td>1.9 MB</td>
</tr>
<tr>
<td>5 mm</td>
<td>73,194,422</td>
<td>2.2 GB</td>
<td>11.0 MB</td>
</tr>
<tr>
<td>10 mm</td>
<td>291,697,949</td>
<td>8.8 GB</td>
<td>23.3 MB</td>
</tr>
<tr>
<td>50 mm</td>
<td>6,853,721,364</td>
<td>205.7 GB</td>
<td>131.1 MB</td>
</tr>
</tbody>
</table>

Table 2-1. Design file size increase according to device diameter and comparison of METAC algorithm.
to-center separation (She et al. 2017a). In our design, we chose the edge-to-edge spacing to be 650 nm considering the fabrication limitations and avoiding the interaction between neighboring antennas. The numerical simulation was carried out using the finite difference time domain method (FDTD module, Lumerical Inc.).

With metasurfaces, the data describing large designs are faced with the challenge of enormous file sizes due to having millions or billions of individual microscopic meta-elements (necessitated by the subwavelength size criterion) described over macroscopically large device areas. This extremely high data density over large areas generates unmanageably large total file sizes, limiting the fabrication of metalenses to sizes no larger than a few millimeters. For example, a 5 cm diameter device may be comprised of over 6 billion meta-elements, each instance of which must be described by nanometer-precision definitions of position and radius, resulting in a size > 200 gigabytes (Table 2-1). Since these design files must subsequently undergo computationally intensive processes such as data conversion (often known as “fracturing”) for use with mask writing equipment, it is critical that these file sizes be minimized.
Figure 2-3. Illustration of an example of data representation of rotationally symmetrically arranged cells at a specific radial position. (a) $S_1$ is the structure of the primitive cell in the library set, which contains the structure to be copied around the circumference of the radial position. In this case, the structure in the cell is a circle of a specified diameter. (b) $S_2$ is the next higher level structure in the library set, which contains a reference $S_1$ as well as a duplicate reference to $S_1$ arranged so that it is adjacent to the first reference of $S_1$. Thus $S_2$ contains twice as many copies as the lower level structure. The base copy angle, $\alpha$, is shown. (c) $S_3$ is the next higher level structure in the library set, which contains two references to $S_2$, and thus 4 copies of $S_1$. (d) Similarly, $S_4$ references $S_3$ twice and contains 8 copies of $S_1$. (e) $S_5$ references $S_4$ twice and contains 16 copies of $S_1$. (f) The top level structure generates the complete device structure at the specified radial position by including the minimum number of unique library references, which includes $S_3$, $S_4$, and $S_5$. 
Figure 2-4. Multiple levels of METAC. Illustration of a possible design of a metasurface lens, where different levels are examined to show the hierarchy of top level structure. The annular structure in Figure 2-3 is actually the 5th ring of this figure, counting from the center. The level numbers indicate which of the referenced structures in the library set are shown with at least this number of levels. For example, the 5th ring only begins to appear when level 3 is shown, corresponding to the reference to the structure S3 in the library set. Levels 4 and 5 show the completion of the 5th ring, corresponding to structures S4 and S5. Only 7 levels are required to complete this device, but in general, larger devices require more levels.
We have implemented a compression algorithm that generates these design files, which we refer to as “METAC” (METAsurface Compression), allowing the file size to be reduced by many orders of magnitude. The algorithm uses a large number of hierarchical levels in the layout file (e.g., GDSII) to copy meta-elements about a central optical axis (most lenses are centrosymmetric), leveraging rotational symmetry to greatly reduce the number of unique meta-element definitions to those along a single radial line. In some cases, certain conversion software or mask writing machines impose an upper limit on the number of levels that any design may contain (e.g., 16), so an algorithm based on METAC but limited to 16 levels, which we call METAC16, was included for reference.

METAC represents data efficiently by making many references to lower level structures (Figure 2-3) in CAD file formats that allow for multiply referenced structures, such as GDSII. For a rotationally symmetric device, at each radial position, a library of self-referenced structures is first generated and then structures within this library are referenced within a top level structure. The library is generated by starting with a primitive structure, S1, which is intended to be copied along an arc corresponding to the circumference of its the radial position, and making a series of exponentially doubled structures with increasing hierarchical levels (Figure 2-4). The primitive structure may be a structure with multiple layers. The second structure in the library, S2, makes two references of S1: one at its original position and a second rotated about the origin of the device such that it is adjacent to the first reference of S1. The third structure, makes two references to S2, and so on. The top level structure, which contains the structure of the intended device, make references to the library by choosing the minimum number of unique structures required from the library to comprise the device at that radial position.
This minimum set of unique structures can be found by converting the total number of copies into a binary number (base-2 numeral system), in which each digit with a value of 1 or 0 corresponds to the inclusion or exclusion of each structure in the library corresponding to the digit number. The structures are then placed at the correct angular positions such that the correct structure is produced. The process is then repeated for each and every radial position until the device is completed. The overview of the procedure is shown in Figure 2-5, and the three main internal steps (calculation of replication parameters, generation of the library set, and creation of references to the library set) are described in Figure 2-6, Figure 2-7, Figure 2-8, respectively.
Figure 2-5. High level procedure for generating layout design file. First, a radial design is determined, in which the geometry of each structure is defined at each radial position. The geometry can be any polygonal shape, and is not limited to circles. Next an empty top level structure is created to contain the final device structure. Finally a loop is performed, where the structure at each radial position is efficiently copied around the arc corresponding to the circumference at the radial position. The copy mechanism is comprised of three main steps: calculation of replication parameters, generation of the library set, and creation of references to the library set, each of which are described in Figure 2-6, Figure 2-7, Figure 2-8, respectively.
Figure 2-6. Calculation of replication parameters. First, the azimuthal angular separation between cells at the same radial position is calculated, which is also referred to as the base copy angle, $\alpha$. This angle can be determined by maximizing the cell density or otherwise defined. Then, $\alpha$ is used to calculate the total number of copies, $N$, including the original, for that radial position. Next, $N$ is converted from a decimal to a binary number, $N_{\text{bin}}$, which is used to determine the references to the library set in a later step, described in Figure 2-8. With that, the calculation of replication parameters is finished for this particular radial position.
Figure 2-7. Generation of library set. First, the primitive structure, S1, is created, which contains the primitive cell, which we would like to be copied. Then a loop is performed, in which increasingly higher level structures are created, each of which contain two references to the next lower level. If the number of levels exceeds the maximum number of allowed levels, which is defined here as $L_{\text{max}} = \text{maximum number of allowed levels} - 2$, then a subroutine is activated in which a structure is created without exceeding the maximum number of allowed levels (2 additional levels are created by this subroutine and also the top level structure). For examples, if there is a limit of 16 levels, then $L_{\text{max}} = 14$. In this way, a library set containing exponentially duplicated copies can be represented.
Chapter 2: Scalable enlargement of metasurfaces for practical purposes

Figure 2-8. Library-set reference creation. Flowchart shows the procedure for creating references to the library set in the top level structure. To efficiently represent the intended structure using the library set, only the minimum number of unique structures in the library set should be included. The unique structures from the library set to be used can be determined by the calculated $N_{bin}$. A loop is performed in which each digit of $N_{bin}$ is queried, such that if the digit is equal to 1, then the structure in the library set with the level number corresponding to the same digit number is referenced in the top level structure; and if the digit is equal to 0, then the reference is not made. Each reference is made at the angular position such that it is adjacent to the previous reference, if it is not the first reference. The first reference may be made at any initial angle, $\gamma$. 

$$\gamma = \gamma + \alpha 2^{M-1}$$

$$M = M + 1$$
Figure 2-9. METAC algorithm performance. (c) Due to the extremely large number of meta-elements required to comprise a large area metalens, a metasurface data compression algorithm, which we call METAC, was developed to generate manageable file sizes of metalens designs. Four methods were compared: uncompressed, EightFold (design is divided into copies of eighths), METAC, and METAC16 (maximum number of levels is restricted to 16, for better compatibility with existing software). File size is plotted as a function of device diameter. (d) The scaling order, $b$, of the file size for each method is plotted. Error bars represent one standard deviation. Methods with $b$ closer to 2 indicate file size scaling with device area, while those with $b$ closer to 1 indicates scaling with device diameter.

We analyzed the METAC algorithm in comparison to other methods of data representation (Figure 2-9), including uncompressed, EightFold, and METAC16. Uncompressed is the method where every meta-element is defined explicitly. Eightfold is the method where the entire design is comprised of eight copies of a single unique octant so that its file size is expectedly one-eighth of the uncompressed. We see that METAC and METAC16 have comparable performance and are extremely effective at compressing file sizes as device diameter increases to the millimeter regime and beyond. We compare a file size scaling order, $b = \log_d F$, which determines the rate at which the file size increases with device dimensions, where $F$ is file size and $d$ is the device diameter.
The uncompressed file size scales with the square of the diameter \((b \approx 2)\), simply because of the element count scales with device area, and similarly with EightFold. It is because of this large \(b\) that metasurface design file sizes quickly reach the gigabyte or even terabyte regime for centimeter scale devices. Ideally, with a metasurface lens, which is greatly centrosymmetric, unique data is contained only along the radial dimension, i.e., ideally \(b = 1\). In fact, METAC yields \(b = 1.073\), with 0.073 attributed to a data definition and reference overhead. Even with METAC16, \(b\) is increased slightly to 1.207, since extra data is required to compensate for the level limit. In the analysis here, the average Voronoi cell length of meta-elements was 520 µm, which is on the large side, indicating that performance gains will be even more pronounced for denser settings. A picture of the output of METAC using GDSII layout file format is shown in Figure 2-10 and a fabricated lens in Figure 2-11. We believe that METAC algorithm will facilitate the development of large area metasurfaces by preventing debilitatingly large file sizes in the realm of terabytes and petabytes as feature densities increase and device areas continue to grow towards the centimeter scale and beyond.
Chapter 2: Scalable enlargement of metasurfaces for practical purposes

Figure 2-10. Metasurface lens section designed using METAC, implemented using using GDSII layout file format.
Chapter 2: Scalable enlargement of metasurfaces for practical purposes

2.3 Large area lenses: fabrication and experiment

Using a design generated by the METAC algorithm, we fabricated metalenses on a 4-inch fused silica (SiO$_2$) wafer substrate (Figure 2-12). A film stack was created over the cleaned substrate, comprised of (from bottom to top) 0.6 µm of a-Si, 1.1 µm of SPR700-1.0 (MEGAPOSIT), and 0.4 µm of CEM365IS (ShinEtsuMicroSi). The a-Si layer was deposited using plasma-enhanced chemical vapor deposition (PECVD by STS). The SPR700-1.0 and CEM365IS were spun-coated at 4000 rpm. The wafer was then

Figure 2-11. SEM of the center of a large metalens designed with METAC. The posts are made of amorphous silicon.
Chapter 2: Scalable enlargement of metasurfaces for practical purposes

Figure 2-12. Metalens production by stepper. Metasurface lenses (metalenses) can be produced at low cost and high yield using existing photolithographic stepper technology. Here a wafer substrate is first deposited with the appropriate film stack, comprised of the metalens material (amorphous (a)-Si), photoresist (SPR700-1.0), and contrast enhancement material (CEM). The pattern of the metalens, which is contained in the reticle, is then projected by the stepper, and replicated rapidly over the face of the wafer by repeatedly exposing and incrementally stepping the wafer position. Throughputs as high as hundreds of wafers per hour (wph) can be achieved. Then the pattern is etched into the a-Si, forming the metalens. Finally, after any residual photoresist is removed, the wafer can be diced into separate individual metalens devices. A photo of fabricated metalens (upper right), 2-cm in diameter, using this methodology is shown in comparison to a ruler. A SEM of the metalens center (center right) shows the microscopic posts comprising the metalens. Scale bar: 2 µm.

exposed using a GCA AS200 i-line stepper. The exposure was followed by a water rinse and post-exposure bake at 115 °C for 60 seconds. The wafer was then developed in MF...
CD-26 (Shipley) for 60 seconds in two baths and immediately rinsed in water. The pattern was then etched into the a-Si using reactive ion etching (STS MPX/LPX ICP RIE). Finally, the resist was stripped by soaking in Remover PG for 10 hours, followed by 2 minutes using the Matrix 105 Plasma Asher (1500 mTorr oxygen pressure, 500 W RF power, 200 °C).

For large area devices, the question of how substrate flatness affects optical performance becomes relevant. Since metasurfaces are usually referred to as phase engineered devices, it is often assumed that strict tolerances must be imposed on phase, and therefore substrate flatness variations must be kept much less than the wavelength over the entire span of the device. However, we have found that tolerances are application dependent, and in particular applications, such as lensing, these tolerances can be relaxed, as opposed to other applications such as holograms and interferograms.

Substrate flatness can be characterized in terms of a surface profile, which are commonly met in various contexts as curvature (warp and bow), total thickness variation, and surface roughness. These properties may be quantified, which exist within a spectrum surface displacement profiles, \( \Delta z(x, y) \), with two parameters: spatial period (or frequency), \( \Lambda \), and perturbation amplitude, \( A \), (Figure 2-13) where surface roughness, warp/bow, and total thickness variation occur with increasing spatial periodicity. Mathematically, we can consider even and odd components, where \( \Delta z = A \cos(2\pi x/\Lambda) \) or \( A \sin(2\pi x/\Lambda) \), respectively, and generally any profile can be represented as a Fourier series of these. The effect of the displacement profile can be understood in terms of both ray and wave optics. In view of ray optics, metasurfaces bend light at the device surface, such that the outgoing angle is determined by the angle of incidence and phase gradient at the
Figure 2-13. Flatness requirements. (a) As metalens size increases, the flatness over the device surface becomes increasingly relevant. The flatness requirements were studied by computer simulations of metalenses (diameter: 2 cm, focal length: 50 mm, design wavelength: 1550 nm), in which the surface curvature was varied (spatial period ranging from 1 to 100 mm, and perturbation amplitude from 0 to 200 µm). The intensity of the optical field is shown. The metalens is situated at the bottom of each plot, and the vertical (z) and horizontal (r) axes are the propagation direction and radial dimension, respectively. Each group four columns on the left and right show the optical behavior for even (cosine) and odd (sine) spatial perturbations, respectively, with respect to the metalens center. Quality of focus is reduced for shorter spatial periods and higher perturbation amplitudes. In (b), the surface profile of the 4-inch wafer we used (including metalenses) was measured (using Toho FLX-2320-S) and the Fourier transform calculated in (c) to obtain the major contributions to spatial frequencies, which mainly occurred at Λ-1 < 0.02 mm-1 (or Λ > 20 mm). The inner and outer white circles denote Λ at 50 and 20 mm, respectively.
length: 50 mm, design wavelength: 1550 nm), we found that the quality of focus worsened as $\Lambda$ decreased and $A$ increased, due to more pronounced angular deviations and interference (Figure 2-13a). As a rule of thumb, good quality of focus (i.e., the substrate is “adequately flat”) is seen for $A \leq 10 \mu$m and $\Lambda > 10$ mm (Figure 2-13a). We measured the surface profile of our wafer as a representative sample, following processing of an array of metalenses, to be $A < 17 \mu$m (Figure 2-13b) and $\Lambda > 20$ mm (Figure 2-13c), satisfying the adequate flatness criteria given by the simulations.

![Figure 2-14. A schematic diagram showing the setup for characterization of the focal spot. A horizontal microscope and a camera are mounted on a motorized stage to scan the image the focal spot, which later can be 3D constructed to determine the focal length and the spot size of a metalens device.](image)

**Experiment**

The focal spot and imaging performance of the metalens were characterized using a horizontal microscope setup (Figure 2-14) and simple imaging setup (Figure 2-15a). A tunable laser (HP 8168F) operating between $\lambda = 1440$-1590 nm with an optical fiber collimator (Thorlabs F810APC-1550) produced a beam 7 mm in diameter, illuminating the metalens center (while the numerical aperture (NA) of the metalens is
Figure 2-15. Focus and imaging performance. (a) The flatness of the device allows for imaging setups very similarly to the ideal thin lens equation, which was used to demonstrate imaging capabilities. (b) Image of focal spot with 7 mm gaussian illumination with $\lambda=1550$ nm. (c) The measured modulation transfer function (MTF) from (b) is plotted with the theoretical diffraction-limited MTF. Error bars: standard deviation. (d) Chromatic focal shift was measured as a function of the wavelength of illumination. The measured deviation of focal length from that of the design wavelength at 1550 nm (light blue dots, error bars: standard deviation) is plotted with the linear fit (blue line). (e) Hyperspectral image of focal spot in the same configuration as (b) for $\lambda=1440-1590$ nm in 10 nm increments linearly binned to RGB channels (center wavelengths $\lambda_R=1590$, $\lambda_G=1515$, and $\lambda_B=1480$ nm). The spot, which is largely white, indicates little chromatic aberration, which can be attributed to the low NA (0.07). Horizontal and vertical line cuts at the RGB center wavelengths are also shown. Using the thin lens setup in (a), simple imaging was demonstrated at $\lambda=1550$ nm for (f) the Harvard university logo and (g) US Air Force 1951 resolution target, without any additional optical components. Scale bar (yellow): 1 mm.
0.2, the limited size of the illuminated area results in an effective NA=0.07). The focal spot was imaged by a horizontal microscope: an objective (10x Mitutoyo M Plan Apo NIR infinity corrected objective), tube lens (Plano-Convex Lens, f = 200.0 mm), and camera (digital InGaAs, Raptor OWL640). The entire horizontal microscope was mounted on a linear motor (NPM Acculine SLP35), which allowed horizontal scanning of the light field. At $\lambda=1550$ nm, the focal length of the device was measured to be 50.159 ± 0.023 mm (design focal length = 50 mm), and the full-width (1/e$^2$) of the focal spot was measured to be 20.9 µm (Figure 2-15b). The theoretical diffraction-limited spot size of this device assuming an aperture of 7 mm and perfect gaussian illumination ($M^2 = 1$) is 14.1 µm. The modulation transfer function (MTF) was calculated by the Fourier transform of the point spread function (i.e., focal spot image) and showed good agreement with the theoretical diffraction-limited MTF (Figure 2-15c).

The focusing efficiency was determined by the ratio of total optical power with and without the metalens measured at the focus position with an optical power meter (Thorlabs PM100D). A pair of irises were used to block any stray light. The efficiency was measured to be 91.8 ± 4.1% at $\lambda=1550$ nm, which does not include losses due to the air-silica boundary (3.25% loss by normal incidence Fresnel reflection). The efficiency may be further improved (e.g. to > 95%) by adding an antireflection coating.

The effect of chromatic aberration was studied in two ways: chromatic focal shift and imaging chromatic aberration, by sweeping the source wavelength while varying and fixing the camera position, respectively. The chromatic focal shift was highly linear ($R^2=0.9982$) and measured to be $\Delta f/\Delta \lambda = -0.0335$ mm/nm, corresponding to a 10% focal shift (-5.09 mm) as the wavelength was tuned by 150 nm (1440-1590 nm) (Figure 2-15d).
A hyperspectral image was taken of the focal spot (Figure 2-15e) to see the imaging chromatic aberration. The focal spot appears white with a faint tinge of color at the periphery, indicating a minor amount of chromatic aberration, which is attributed to the low NA. This is corroborated by the horizontal and vertical line cuts, which coincide closely.

The setup was modified to perform direct single lens imaging of macroscopic objects (Figure 2-15a). Due to the thinness of the metalens, this configuration closely approximated that of the traditional thin lens equation (Hecht 2002), which is commonly taught in introductory optics textbooks yet does not accurately describe the imaging conditions of real, bulky lenses, since it assumes a lens of negligible thickness. The entire setup consisted of an illuminated object, the metalens, and camera. Objects were patterns made using chromium on glass, in the form of the Harvard University logo and USAF 1951 resolution target (Figure 2-15f,g respectively). The images were brought into focus by adjusting the object-to-lens distance \(d_o\) to 175.0 ± 0.5 mm, and lens-to-camera distance \(d_i\) to 70.0 ± 0.5 mm. This agrees with the thin lens equation \(1/d_i + 1/d_o = 1/f\), where and \(f\) is the focal length) for a lens with \(f = 50\) mm. The optical magnification was measured to be -0.42, since a 4 mm reference object was inverted and imaged to have a length of 1.69 mm on the camera sensor. The corresponding calculated value according to the thin lens magnification formula, \(M = f/(f - d_o) = -0.40 ± 0.003\).
2.4 Outlook: lenses of the future

We have shown the design, fabrication, and optical characterization of the large area metalenses, 2 centimeters in diameter with efficiency greater than 91%, fabricated using photolithographic steppers. Our algorithm for large area, high data density metasurface designs is general: in addition to the photolithographic method described, the algorithm can be used for other fabrication methods, such as nanoimprint and self-assembly based techniques (Chen et al. 2015; O’Hern et al. 2015; Moitra et al. 2015).

The metalens, which is only 600 nm thick, in combination with the 0.5 mm thick substrate is a close experimental approximation of a lens described by the ideal thin lens equation. Optical components produced using this method have several advantages. In contrast to machining, such as diamond turning (Saito 1975) and magnetorheological finishing (Kordonski et al. 2004), which form the shape of lenses very precisely but relatively slowly, photolithography can imprint properties of an optical component with a flash exposure only milliseconds in duration (Mack 2007). Also, in contrast to molds (Schaub 2011), a projection photomask can be used an unlimited number of times without any wear or loss of fidelity. Finally, fabrication processes are shouldered by mature semiconductor manufacturing technology.

Moore’s law, which describes the trend where the number of transistors per square inch doubles every year (Moore 1965), is mainly defined, among other factors, by the rate at which lithographic techniques can shrink down pattern feature sizes, i.e. technology nodes sizes (Mack 2015; Bruning 2007; Wikichip.org 2017) (Figure 2-16).
1992, the technology node size dropped below the 700 nm, which is coincidentally the red end of the visible spectrum (Fletcher 1995). From that point forward, the mass production of subwavelength metasurfaces working at visible wavelengths became possible. Using high-yield technologies such as photolithographic steppers and scanners (Fay 2002), we envision a manufacturing transition from using machined or molded optics to lithographically patterned optics, where they can be mass produced with the similar scale and precision as IC chips in semiconductor fabrication facilities, such as foundries. An arrangement, in which optical and electronic parts are produced in the same place with the same equipment, is also naturally advantageous for integration. State-of-the-art equipment is useful, but not necessarily required. In particular, as it becomes ever more difficult to keep pace with Moore’s law at the time of writing and as state-of-the-art lithography systems, such as extreme ultraviolet lithography (EUVL) (Bergmann et al. 2010), become so overwhelmingly expensive such that they are inaccessible to everyone except for the largest players in the industry, it is important to find avenues in which existing capital equipment can be brought out of obsolescence and repurposed for new, exciting opportunities.
Figure 2-16. Moore’s Law and metasurfaces. Being an early enabler of metasurfaces, Moore’s law, which predicts the transistor areal density in computer chips to double every year, is driven in large by improvements in lithographic technology. The plot shows the state of lithographic technology, represented as technology node size (TNS, diamond symbols), as a function of year, where smaller TNSs indicate higher feature densities. Several important product landmarks utilizing these TNSs are labelled, and key leaps in TNS powered by new light sources are denoted by vertical stems with circles. These past developments will enable the possibility of large area, cost-efficient metasurfaces optical devices. The red dotted line denotes the 700 nm threshold (corresponding to red light), below which the TNS had already surpassed in the mid-1990s. The wavelengths of the visible spectrum corresponding to TNS are shown shaded from 400-700 nm. In general, subwavelength-sized features can be produced using TNSs with at least twice the feature density as compared to the wavelength of light. Subsequent improvements to TNS up to the present (blue shaded region) have enabled feature sizes much smaller than the wavelength of light, providing ways to create more complex, fine-featured meta-elements as well as alternative routes for effectively utilizing foundry equipment which would otherwise be viewed as obsolete.
Chapter 3
Dynamic wavefront shaping with adaptive metalenses

3.1 Introduction

Focal adjustment and zooming are universal features of cameras and advanced optical systems. Such tuning is usually performed longitudinally along the optical axis by mechanical or electrical control of the focal length. However, the recent advent of ultrathin planar lenses based on metasurfaces (metalenses), which opens the door to future drastic miniaturization of mobile devices such as cell phones and wearable displays, mandates fundamentally different forms of tuning based on lateral rather than longitudinal motion. Theory shows that the strain field of the metalens substrate can be directly mapped into the outgoing optical wavefront to achieve large diffraction-limited focal length tuning and control of aberrations. We demonstrate experimentally electrically-tunable large area metalenses controlled by artificial muscles, which are capable of simultaneously performing focal length tuning (>100%) as well as on-the-fly astigmatism and image shift corrections, which until now were only possible in electron optics. The device thickness is only 30 microns. Our results demonstrate the possibility of future optical microscopes, which operate fully electronically, as well as compact
optical systems that employ the principles of adaptive optics to correct many orders of aberrations simultaneously.

There has been a wide variety of work on tunable optical devices, in particular, tunable focus lenses (i.e., varifocal lenses), with important applications in imaging and adaptive vision. Focus tuning is usually performed by the mechanical movement of rigid elements (Kasunic 2012), which provides great imaging performance at the expense of bulk and inertially-limited speed. Other implementations include liquid crystal based spatial light modulators (Itoh et al. 2009; Li, Bryant, Van Heugten, et al. 2013; Li, Bryant, van Heugten, et al. 2013), which can be relatively high speed and capable of correcting for aberrations, but are limited in resolution and polarization dependence; fluid based tunable lenses (Shian et al. 2013; Hasan et al. 2017; Maffli et al. 2015; Dong et al. 2006; Olles et al. 2011; Ren & Wu 2007) (including those incorporating elastomeric materials), which can be high speed, have wide tuning range, and are relevant for small-scale devices, but difficult to control the exact surface curvature (e.g. coma has been observed when placing the lens vertically, due to gravity); tunable acoustic gradient index lenses (Mermillod-Blondin et al. 2008), which can be tuned at high speeds but must be used stroboscopically; and electro-optic lenses (Shibaguchi & Funato 1992), which are high speed but polarization-sensitive. In astronomy (Babcock n.d.; Ragazzoni et al. 2000) and microscopy (Ji 2017), adaptive optics use deformable mirrors to correct for wavefront distortions, but must operate in reflection.

Recently, flat lens technology based on metasurfaces (Yu et al. 2011), which control the wavefront of light using subwavelength-spaced nanostructures, has shown considerable potential in optical performance while reducing the element thickness to the
micron level, opening up new opportunities to replace bulk optical devices, with thin, flat, lightweight devices (Yu & Capasso 2014; Genevet et al. 2017). Instead of moving several optical components longitudinally along the optic axis as in telephoto lenses and autofocusing cameras, in metasurfaces lateral control can be utilized for varying focus and magnification, as well as other adaptive optics capabilities (Babcock n.d.; Ragazzoni et al. 2000; Ji 2017), leverage their flatness. Mechanically-tunable metasurfaces have been reported recently, in which metasurfaces are embedded in stretchable substrates (Ee & Agarwal 2016; Malek et al. 2017; Kamali et al. 2016), but were limited in size (less than a millimeter in diameter), required external apparatuses to apply strain, and had inherent speed limitations, restricting their applications. In addition, no theoretical analysis of the effect of tuning the strain field on optical aberrations was presented. Here the combination of metasurface optics and dielectric elastomer actuators (DEAs), sometimes known as artificial muscles (O’Halloran et al. 2008) in soft robotics, offers a versatile platform for electrically-tunable optical devices, through the design of phase, amplitude, and polarization profiles.

### 3.2 Theory of strain-field mediated tuning

A flat lens can be constructed by a metasurface with the following hyperboloidal phase profile (Wang et al. 2017):

\[
\varphi(r) = \pm k \left( \sqrt{r^2 + f^2} - f \right)
\]

(3.1)

where \( k \) is the wavenumber, \( r \) is the radial position, \( f \) is focal length, and the sign is positive or negative for diverging or converging lenses, respectively (Figure 3-1). A lens
Figure 3-1. Principle of strain field-mediated tunable metalens. A metasurface (left column) is constructed by digitizing an analog optical phase profile on a flat surface into discrete cells, each of which contains a metasurface element that locally imparts the required phase shift to the incident light in order to reconstruct the desired wavefront (middle column, dotted line: optic axis). The wavefront generated by the metasurface determines the subsequent beam shaping (right column). (a) Original: metasurface without stretch. (b) Defocus: metasurface with uniform and isotropic stretch. (c) Astigmatism: metasurface under asymmetric stretch. (d) Shift: metasurface displaced laterally in the x,y plane.

described by Equation 1 focuses light free of spherical aberrations for normal incidence (infinity-corrected) illumination. By combining this metasurface with a DEA, the lens can be made electrically tunable.
A DEA (O’Halloran et al. 2008), a type of electroactive polymer, is effectively a compliant parallel plate capacitor, which stretches when a voltage or electric field is applied (Brochu & Pei 2010). By using a soft elastomer (e.g., polyacrylate and silicone rubbers) as a dielectric together with transparent, stretchable electrodes (Shian et al. 2013), the configuration is slightly compressed in the thickness direction when a voltage is applied. As elastomers conserve volume on deformation, the thinning results in a lateral expansion that can be very large, ~500% (Huang et al. 2012). The attainable actuation strain is a function of the electrostatically-induced Maxwell stress, the constitutive deformation behavior of the elastomer, and the mechanical configuration of the elastomers and electrodes (Koh et al. 2011). Bonding a metasurface with the DEA couples the metasurface profile to the voltage-induced stretching, which is approximately uniform over a single electrode area. Effectively, the coordinates of the metasurface scale by a stretch factor, $s$ ($s = 1 + \varepsilon_L$, where $\varepsilon_L$ is lateral strain), such that $\mathbf{r} \rightarrow \mathbf{r}/s$:

$$q_{\text{stretched}}(r,s) = \pm k \left( \sqrt{(r/s)^2 + f_0^2} - f_0 \right),$$

(3.2)

where $f_0$ is the focal length without actuation (Figure 3-1b). The focal length of Equation 3.2 varies quadratically with stretch: $f(s) = s^2 f_0$. The voltage-dependence of focal length provided by the DEA is:

$$\frac{f}{f_0} = \frac{1}{1 - (\varepsilon/Y)(V/t)},$$

(3.3)
Chapter 3: Dynamic wavefront shaping with adaptive metalenses

<table>
<thead>
<tr>
<th>Noll index (j)</th>
<th>Classical name</th>
<th>Radial degree (n)</th>
<th>Azimuthal degree (m)</th>
<th>Zernike term ((Z_n^{(2)}) or (Z_{n}^{(3)}))</th>
<th>Uniformly Stretched Zernike Term ((Z_j^{(2)} = Z_j(r \rightarrow r/s)))</th>
<th>Re-scalable? (Y/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Piston</td>
<td>0 0</td>
<td></td>
<td>(1)</td>
<td>(1)</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>Tip (X-Tilt)</td>
<td>1 1</td>
<td></td>
<td>(2r \cos(\theta))</td>
<td>(2r \cos(\theta)/s)</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>Tilt (Y-Tilt)</td>
<td>1 -1</td>
<td></td>
<td>(2r \sin(\theta))</td>
<td>(2r \sin(\theta)/s)</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>Defocus</td>
<td>2 0</td>
<td></td>
<td>(\sqrt{3}(2r^2 - 1))</td>
<td>(\sqrt{3}(2r^2/s^2 - 1))</td>
<td>Y</td>
</tr>
<tr>
<td>5</td>
<td>Oblique</td>
<td>2 -2</td>
<td></td>
<td>(2\sqrt{6}r^2 \sin(2\theta))</td>
<td>(2\sqrt{6}r^2 \sin(2\theta)/s^2)</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>Vertical</td>
<td>2 2</td>
<td></td>
<td>(2\sqrt{6}r^2 \cos(2\theta))</td>
<td>(2\sqrt{6}r^2 \cos(2\theta)/s^2)</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>Vertical coma</td>
<td>3 1</td>
<td></td>
<td>(\sqrt{3}(3r^3 - 2r) \sin \theta)</td>
<td>(\sqrt{3}(3r^3/s^3 - 2r/s) \sin \theta)</td>
<td>N</td>
</tr>
<tr>
<td>8</td>
<td>Horizontal coma</td>
<td>3 1</td>
<td></td>
<td>(\sqrt{3}(3r^3 - 2r) \cos \theta)</td>
<td>(\sqrt{3}(3r^3/s^3 - 2r/s) \cos \theta)</td>
<td>N</td>
</tr>
<tr>
<td>9</td>
<td>Vertical</td>
<td>3 -3</td>
<td></td>
<td>(\sqrt{6r^3} \sin(3\theta))</td>
<td>(\sqrt{6r^3} \sin(3\theta)/s^3)</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>Oblique</td>
<td>3 3</td>
<td></td>
<td>(\sqrt{6r^3} \cos(3\theta))</td>
<td>(\sqrt{6r^3} \cos(3\theta)/s^3)</td>
<td>Y</td>
</tr>
<tr>
<td>11</td>
<td>Primary</td>
<td>4 0</td>
<td></td>
<td>(\sqrt{5}(6r^4 - 6r^2 + 1))</td>
<td>(\sqrt{5}(6r^4/s^4 - 6r^2/s^2 + 1))</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>Vertical</td>
<td>4 2</td>
<td></td>
<td>(\sqrt{10}(4r^4 - 3r^2) \cos(2\theta))</td>
<td>(\sqrt{10}(4r^4/s^4 - 3r^2/s^2) \cos(2\theta))</td>
<td>N</td>
</tr>
<tr>
<td>13</td>
<td>Oblique</td>
<td>4 -2</td>
<td></td>
<td>(\sqrt{10}(4r^4 - 3r^2) \sin(2\theta))</td>
<td>(\sqrt{10}(4r^4/s^4 - 3r^2/s^2) \sin(2\theta))</td>
<td>N</td>
</tr>
<tr>
<td>14</td>
<td>Vertical</td>
<td>4 4</td>
<td></td>
<td>(2\sqrt{10}r^4 \cos(4\theta))</td>
<td>(2\sqrt{10}r^4 \cos(4\theta)/s^4)</td>
<td>Y</td>
</tr>
<tr>
<td>15</td>
<td>Oblique</td>
<td>4 -4</td>
<td></td>
<td>(2\sqrt{10}r^4 \sin(4\theta))</td>
<td>(2\sqrt{10}r^4 \sin(4\theta)/s^4)</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 3-1. Wavefront shaping with the Zernike polynomials. A series of Zernike terms can express any wavefront function or metasurface phase profile over a unit disk: 
\(\phi(r, \theta) = \sum_j a_j Z_j(r, \theta)\), where \(j\) is the Noll index and \(a_j\) is the coefficient amplitude of \(Z_j\). Also shown are the Zernike terms after uniform stretching, in which the radial coordinate transforms as \(r \rightarrow r/s\). The stretched Zernike term is indicated as “re-scalable” if the \(s\) dependence can be factored out cleanly (while ignoring any constant offset), effectively yielding a tunable coefficient of \(Z_j\), in which the form of \(Z_j\) is preserved when \(s\) is varied, thus indicating that the particular term is a good candidate to be tuned via uniform stretching.

45
where $V$ is the voltage, and $\varepsilon$, $Y$, and $t$ are the permittivity, Young’s modulus, and dielectric layer thickness, respectively. Tuning in this manner introduces negligible aberrations (Figure 3-2).

In practice, an infinite number of wavefront aberrations can exist and can be quantified in terms of Zernike polynomials (Table 3-1). In most cases, the lowest eight terms are sufficient: piston, tip, tilt, defocus, oblique and vertical astigmatism, and vertical and horizontal coma. Because the Zernike terms are linear and orthogonal, specific aberrations can be targeted and tuned by introducing the appropriate displacement field to the phase profile. In general, applying stress of a particular configuration induces a strain field, resulting in a displacement field. The displacement field can be regarded as the sum of deformation ($\vec{A}$) and a rigid body displacement ($\vec{B}$) components, in which the transformed coordinates can be expressed in the following way: $\vec{x}' = \vec{A}(\vec{x}, \eta) + \vec{B}(t)$, where $\eta$ is a parameter (e.g., voltage) and $\vec{x}$ and $\vec{x}'$ are the original and transformed coordinates, respectively (Timoshenko & Goodier 1983). For metasurfaces, the phase profile transform as $\phi(x, y) \rightarrow \phi(x', y')$. The rigid body displacement is simply a lateral shift of the entire phase profile, while the deformation changes the shape or size, which results in a change of shape to the passing wavefront. In the case of asymmetric biaxial strain, the coordinates of the phase profile transform as $(x, y) \rightarrow (x/s_x, y/s_y): \phi = \pm k \left( \sqrt{\left(\frac{x}{s_x}\right)^2 + \left(\frac{y}{s_y}\right)^2} + f'_0 - f_0 \right)$, where $s_x \neq s_y$. Light propagating along two perpendicular planes experience different focal lengths, which is astigmatism. Thus through electrically controlled strain field, it is possible to create the optical analog of image shift and stigmators (Figure 3-1c,d) found in electron microscopes (Wischnitzer 1981).
Figure 3-2. Uniform stretch approximation. (a) The ideal stretch profile \( s_{\text{ideal}} \), which is the stretch profile that produces an aberration-free hyperboloidal phase profile) and the nominal, uniform stretch value \( s \) is plotted for typical values of \( s \) from 1.01 to 1.3 (they are the same for \( s = 1 \)) as a function of radial position. (b) The difference, \( \Delta s = s_{\text{ideal}} - s \), is plotted for typical values of \( s \) from 1 to 1.3 (blue to red lines, respectively), as a function of radial position for an example lens with focal length of 50 mm. (c) The magnitude of the wavefront aberration function (WAF) in radians (color bar), here defined as the difference between stretched \( \varphi_{\text{stretched}} \) and ideal \( \varphi_{\text{ideal}} \) phase profiles, is plotted as a function of radial position \( r \) and stretch \( s \) for a lens with focal length of 50 mm using light of wavelength 1550 nm. The white, dotted line shows the maximum aberration, which occurs at a stretch of \( s \approx 1.22 \).
Uniform stretching approximation

A lens that focuses normal incidence light has the following hyperboloidal phase profile:
\[ \phi_{\text{ideal}} = \pm k \left( \sqrt{r^2 + (s^2 f_0)^2} - s^2 f_0 \right), \]
where we desire a tunable focal length relationship of the form \( f = s^2 f_0 \). In order to achieve \( \phi_{\text{ideal}} \), a stretch profile that depends on the radial position is required: \( s_{\text{radial}}(r) \). Then the phase profile of a lens with focal length \( f_0 \) that is stretched by \( s_{\text{radial}} \) undergoes the coordinate transformation: \( r \rightarrow r / s_{\text{radial}}(r) \), such that the resulting phase profile is the following:
\[ \phi_{\text{radial}} = \pm k \left( \sqrt{(r / s_{\text{radial}}(r))^2 + f_0^2} - f_0 \right). \]

In order to find the stretch profile that produces the target phase profile \( \phi_{\text{ideal}} \), which has a nominal, uniform stretch value \( s \), we can equate \( \phi_{\text{radial}} = \phi_{\text{ideal}} \) and write \( s_{\text{ideal}} \) in place of \( s_{\text{radial}} \) (where \( s_{\text{ideal}} \) is a specific case of \( s_{\text{radial}} \) that produces \( \phi_{\text{ideal}} \)):
\[ \sqrt{(r / s_{\text{ideal}}(r))^2 + f_0^2} - f_0 = \sqrt{r^2 + (s^2 f_0)^2} - s^2 f_0. \]
Solving for \( s_{\text{ideal}} \), we find the following:
\[ s_{\text{ideal}}(r) = r / \sqrt{r^2 + 2(s^4 - s^2) f_0^2 + 2(1 - s^2) f_0 \sqrt{r^2 + (s^2 f_0)^2}}. \]
To lowest order, we see that \( s_{\text{ideal}} \approx s \).

Although \( s_{\text{ideal}} \) is not exactly equal to \( s \), their difference is extremely small. Figure 3-2a,b show that there is a slight decrease in local stretch as the radial position is moved outwards from the center of the lens, using the lens parameters of our experiment (\( f_0 = 50 \) mm). A nominal stretch of \( s=1.2 \) yields a difference of only \( \Delta s = 1 \times 10^{-4} \) at a radial position of 3 mm, corresponding to the size of the lens in our device.
Focal length vs. voltage relation

Volume conservation of the elastomer requires that: \( s_x s_y s_z = 1 \), where \( s \) is the stretch. For isotropic materials under uniaxial compression: \( s_x = s_y = s \), where \( s = 1 + \epsilon_L \) (\( \epsilon_L \) is the lateral strain) and \( s_z = 1 + \epsilon_z \) (\( \epsilon_z \) is the longitudinal strain, i.e. in the thickness direction), giving \( s^2 = 1/s_z \). For small strain (\( \epsilon \ll 0.2 \)), the strain response \( \epsilon_z \) of materials under Maxwell’s compression stress, \( \sigma_M = -\varepsilon E^2 \), can be approximated as that of a linear elastic material with Young’s Modulus of \( Y \): \( \epsilon_z = -\varepsilon E^2 / Y \), where \( E \) is the electric field and \( \varepsilon \) is permittivity. As \( s_z = 1 + \epsilon_z \) and \( f/f_0 = s^2 \), the relation of focal length with voltage, \( V \), is the following (Equation 3.3): \( \frac{f}{f_0} = \frac{1}{1 - (\epsilon / Y)(V/t)^2} \), where \( t \) is the instantaneous thickness of the elastomer, which is a function of the applied voltage, but can be approximated as constant since it changes very little (Pelrine 2000). By writing \( b = \epsilon / (Y t^2) \), the series expansion for small voltages near \( V = 0 \) gives \( f/f_0 = 1 / (1 - b V^2) = 1 + b V^2 + b^2 V^4 + O(V^6) \). We see that the focal length relation transitions from a quadratic relation \( (V^2) \) to a quartic relation \( (V^4) \), as the voltage is increased, \( V_{\text{transition}} > b^{1/2} \) (and to higher orders if \( V \) is increased further). For example, by taking nominal values for our dielectric layer made of VHB: \( \epsilon_r = 6, Y = 1.8 \text{ MPa} \), and \( t = 30 \mu\text{m} \), \( V_{\text{transition}} \) is 5.52 kV.
Aberrations introduced by stretching

In tunable systems, optical aberrations may be augmented or reduced by tuning. With metalenses, spherical aberration using normal incidence illumination is corrected by design. However, when such a metasurface is deformed, the resulting aberration is not obvious. The wavefront aberration function \( \text{WAF} = \phi_{\text{stretched}} - \phi_{\text{ideal}} \) quantifies the deviation of the resulting phase profile of the stretched metasurface from the ideal phase profile:

\[
\text{WAF} = \pm k \left( \sqrt{\frac{r}{s}} \frac{f_0}{f} - f_0 \right) \pm k \left( \sqrt{\left(1 - s^2\right)\frac{r^4}{8}} + O(s^6) \right).
\]

This equation shows that as the uniform stretch is increased from \( s = 1 \), the WAF increases from 0 until reaching a maximum aberration (Figure 3-2c) at \( s \approx 1.22 \) (e.g., for a lens \( \varnothing 6 \text{ mm}, f = 50 \text{ mm}, \) the maximum aberration is \( <0.05 \) rad at the edge). Upon further stretching, a built-in suppression of spherical aberration comes into effect, in which the WAF decays following a quartic function, \( (r/s)^4 \). This allows for highly tunable lens devices with excellent immunity to aberration.

3.3 Device design, fabrication, and assembly

A polarization-insensitive, converging metalens (diameter \( \varnothing 6 \text{ mm}, f = 50 \text{ mm}, \) \( \lambda = 1550 \text{ nm} \)) was combined with a DEA constructed using transparent polyacrylate elastomers with stretchable-transparent patterned electrodes made of single-walled carbon nanotubes (SWCNT) (Shian et al. 2012). The DEA was measured to exhibit large transparency windows in the visible, near-infrared, and mid-infrared spectra (Figure 3-3). Focal length tuning was implemented by applying a voltage through the center electrode
Figure 3-3. Optical measurements of DEA, consisting of pre-stretched (4x) acrylic elastomer (VHB 4905, 3M) and single-walled carbon nanotube (SWCNT) electrodes. The measurements were performed using the Agilent Cary 7000 Universal Measurement Spectrometer and the Bruker Lumos FTIR microscope. The spectrum range is from 200 nm to 16 µm. The x-axis of the plots is in log scale. Our device operates at 1550 nm. Top: Transmission (blue), reflection (green), and absorption (yellow) measurements for a single VHB membrane layer. Middle: Transmission, reflection and absorption measurements for a device with the VHB layer and single-walled carbon nanotube (SWCNT) electrodes applied to both sides. Bottom: Derived absorption spectrum of SWCNT layers only. At 1550 nm, the absorption is close to zero.
Figure 3-4. Dielectric elastomer actuator (DEA) metalens device design. (a) A schematic of the device in which a metalens and DEA with 5 addressable electrodes are combined to allow for electrical control over the strain field of the metasurface. (b) Optical microscope images (scale bar: 20 µm) at (i) no voltage, (ii) 2.5 kV applied to the center electrode (V5), and (iii) 2.75 kV applied to tune X-astigmatism (concurrently, V1 and V3). The dark spots are defects (either missing or tilted silicon posts) introduced during the transfer process. The corresponding 2D Fourier transforms of the images (i)-(iii) are shown in (iv)-(vi), respectively (normalized amplitudes).

V5 for increasing focal length or from V1 to V4 together for decreasing focal length (Figure 3-4), corresponding to lateral expansion or contraction of the post spacings, respectively. The control of vertical astigmatism in the x, y directions (“x,y-stigmators”) was implemented by activating opposing pairs of electrodes. Image shift was implemented by activating one peripheral electrode: V1 through V4, such that its expansion causes the entire metasurface to shift in space. Any combination of actuations, each implemented using different voltages, is also possible.

Two types of devices were fabricated and measured, a single layer (SL) and double layer (DL) device. Although the SL demonstrated better tunability than the DL (for the same applied voltage, the SL was stretched more than DL, on account of the lesser stiffness introduced by the smaller intermediate elastomer layer), most data
presented are of the DL, due to the device being higher quality. The maximum voltage used was 3 kV, producing $s = 1.41$ (SL) and 1.15 (DL).

Figure 3-5. Metalens design elements. (a) The unit cell of the lens consisting of the SiO2 substrate and an amorphous silicon (a-Si) post. The dimension are labeled as follows: $h$ is the height of the post, $d$ is the diameter of the post, $u$ is the size of the unit cell and $e$ is the constant edge-to-edge spacing between neighboring posts. (b) Phase and amplitude response of posts are plotted as a function post diameter. (c) The metasurface design of our device is shown. Left: The whole lens is 2 cm in diameter and appears as a solid black circle in this rendering due to the sheer density of structures. Top right: a close up view showing a 200 µm wide window of the center of the lens. Bottom right: a close up view showing a 50 µm wide window of the center of the lens.
Metalens design

The unit cell of the design is shown in Figure 3-5. The height of the posts is $h = 950$ nm. By varying the diameter of the posts ($d = 810 - 990$ nm), a phase coverage of close to $2\pi$ and a high, uniform transmission amplitude response are achieved. These data were used as a lookup table to digitize the phase profile. Post diameters with low transmission values (e.g., $d = 860$ and 870 nm) were excluded from the lookup table. Given the circular shape of the post structures, the phase and amplitude responses are independent of the polarization of the incoming light. The phase profile was realized by subwavelength antennas with fixed edge-to-edge separation, by which the placement of antennas is made denser than that with the conventional fixed center-to-center separation. Hence the size of the unit cell ($u$) is equal to the sum of the post diameter ($d$) and the constant edge-to-edge spacing ($e$): $u = d + e$. In our design, we chose the edge-to-edge spacing to be 650 nm, which was determined by the feature size of the stepper we used as well as the length scale to avoid interaction between neighboring antennas.
Chapter 3: Dynamic wavefront shaping with adaptive metalenses

Selection of materials

A wide range of elastomers with varying chemistry, mechanical, and dielectric properties have been reported for electrostatic actuation. We chose an off-the-shelf acrylate elastomer (VHB 4905, 3M) because it is transparent, sticky, easily available, and capable of large strain (Pelrine 2000). Although VHB is limited in choice of thickness and not purposely optimized for DEA applications, the use of VHB provide a balance between ease of preparation and optical/actuation performance. Silicone-based elastomers, which are also transparent, offer more precise control, larger temperature stability, and very low hysteresis, but their use requires additional processing steps.

Transparent, stretchable electrodes are prepared from single-walled carbon nanotubes (SWCNTs). A controlled, uniform distribution of SWCNT mats is achieved using a water-based SWCNT dispersion and filtration method. This method produces SWCNT mats on a PTFE filter that can be directly transferred onto elastomers by pressing them and patterning them with a mask. Applications of DEAs for in-line optical device require optimal density of SWCNTs (Huang et al. 2012). A low density is desired to minimize optical absorption and mechanical stiffening of the actuator. However, a higher density is needed to maintain electrical percolation (particularly for large-areal actuation) and to minimize the RC constant.

In order to transfer a metasurface from its substrate to the elastomer membrane, a sacrificial layer was used. It is important for the sacrificial layer to be soluble in a solvent orthogonal to both the metasurface and the elastomer, such that the solvent only dissolves the sacrificial layer but not the metasurface nor elastomer. We chose germanium dioxide (GeO$_2$) as the sacrificial layer, since it can be readily dissolved in water. Water enables
versatility of the process to be used with a wide variety of metasurface materials and membrane materials, as well as in other applications. We also tried using polyacrylic acid (PAA) as a polymer-based sacrificial layer, which is soluble in water, but did not obtain good results due to incompatible high temperature processing steps (Figure 3-6).

Figure 3-6. Polyacrylic acid (PAA) based sacrificial layer. PAA was attempted to be used as the water-soluble sacrificial layer. Metalenses are shown with damaged film on a hot plate. High temperature processing steps as well as exposure to humidity and water caused difficulty in the process (in particular, from thermal expansion of the polymer).
Device fabrication

A film stack was first prepared for nanofabrication (Figure 3-7). Starting with a silicon wafer, a 0.4 µm layer of elemental germanium (Ge) was deposited by electron beam evaporation. The sample was then placed in a furnace (Tystar Tytan) for dry oxidization in the presence of O₂ at atmospheric pressures at 550 °C for 3 hours, which converted the layer of Ge into GeO₂, completing the sacrificial layer. We found that the GeO₂ layer made by dry oxidizing Ge was significantly more soluble in water than layers deposited by either electron evaporation or thermal evaporation directly using GeO₂ as the source material, possibly due to formation of GeOₓ species at high temperatures. GeO₂ also dissolves much faster in water than Ge in hydrogen peroxide (H₂O₂) and water solution, in which the rate-limiting step is the oxidation of Ge into GeO₂ by H₂O₂. The metasurface itself is composed of nano-posts made of amorphous silicon (a-Si) with height of 950 nm, so a layer of a-Si with the corresponding thickness was deposited using plasma-enhanced chemical vapor deposition (PECVD).

In preparation for photolithography, the sample surface was first spin coated with the adhesion promoter, hexamethyldisilazane (HMDS), at 4000 rpm. Next, a 1 um layer of i-line photoresist (SPR700-1.0, DOW) was spun-coated and soft baked at 95 °C for 60 s. Over the photoresist, a layer of photo-bleachable contrast enhancement material (CEM365iS, ShinEtsuMicroSi) was spun-coated at 4000 rpm to improve feature contrast for the following stepper exposure.
Figure 3-7. Dielectric elastomer actuator platform. (a) Fabrication steps: (i) Schematic of the thin-film stack prepared for photolithography. (ii) After patterning and etching, the stack is attached to an elastomer layer and the whole stack is immersed in a water bath. (iii) Schematic of the release process showing the dissolution of the sacrificial layer (GeO$_2$) from the outer edge of the device towards the center, leaving the metasurface structure attached to, and supported by, the elastomer layer. (iv) A schematic of metasurface and DEA combination in which an applied voltage supplies the electrode layers (SWCNTs) with electrical charges. Their electrostatic attraction compresses the elastomer in the thickness direction, and causes expansion in the lateral direction. (b) False-colored scanning electron microscope image of the center of the lens shows a-Si posts on the GeO$_2$ sacrificial layer before attaching to the elastomer layer.
Figure 3-8. Dissolution of a GeO$_2$ based sacrificial layer in water, releasing a thin film of silicon.

The metasurface design was patterned into the photoresist using stepper photolithography, which allowed us to produce large area metalenses with high yield. A quartz photomask was patterned using a high resolution laser lithography system (Heidelberg DWL2000). This photomask was then used as the reticle in a 5x reduction i-line stepper (GCA AS200 AutoStep) to expose the photoresist. After exposure, the CEM365iS layer was removed by spraying deionized water (DIW) and spin drying. The sample was then post exposure baked (PEB) at 115 °C for 60 s. The photoresist was then soaked in developer (MF CD26) for 90 s and rinsed in DIW. A mild O$_2$ plasma descum was performed to improve pattern fidelity.

The photoresist pattern was used as the etch mask to create the metasurface in the a-Si layer. The sample was etched using an inductively coupled plasma (ICP) reactive ion
etch (RIE) system (STS MPX/LPX ICP RIE), in which the etchant gases used were perfluorocyclobutane (C₄F₈) and sulfur hexafluoride (SF₆). The pattern was etched completely through the a-Si layer and as much as 100 nm into the sacrificial layer.

Finally, the photoresist was removed by soaking the sample in N-methyl-2-pyrrolidone (NMP) solution (Remover PG, Microchem) for 8 hours followed by a dry resist strip in a high temperature (200 °C) and high power (500 W) O₂ plasma asher (Matrix Plasma Asher) for 20 s, leaving a-Si and GeO₂ exposed.

To create the DEA, a membrane of elastomer, VHB, was uniformly biaxially stretched (4x, linear) and mounted by its own adhesion on a rigid circular plastic ring. SWCNT electrodes were then applied to either side of the membrane (~30 µm thick). The simplest electrode configuration (single area) is a disk of SWCNTs concentrically applied to both sides of the membrane in equal areas, so as to produce a single uniform deformation region. More complex electrode configurations were made by patternning multiple electrodes on the same membrane, for instance, by putting multiple addressable patches one side of the membrane and a common ground electrode on the other side. We fabricated two kinds of electrode patterns: (1) a single area and (2) five-segment area. The five segment device allows for the creation and control of different strain fields, including radial expansion and contraction strains in the center, uniaxial strain in both the x and y directions, as well as rigid body displacements in both directions.
Chapter 3: Dynamic wavefront shaping with adaptive metalenses

Figure 3-9. Device design and operation. (a) The device with the metasurface at the center of a pre-stretched DEA sheet, suspended by a plastic frame. The electrical contacts are made to the SWCNT electrodes using silver grease and carbon tape (black stripes). There are eight connections using SWCNT, labeled $V_1$-$V_5$ and GND (ground). Schematics (cross section) of single layer (SL) and double layer (DL) devices are shown. Both devices have the same basic structure in which the metasurface is attached to a DEA via an intermediate elastomer layer (IEL). (b) Different electrode activation configurations. Applying a voltage to an electrode will cause the area under that electrode to expand. The defocus (+) tuning is controlled by the center electrode, while (-) tuning is controlled by applying voltages of equal magnitude to four peripheral electrodes. Astigmatism (X or Y) is controlled by opposing pairs of peripheral electrodes. The X, Y shift is controlled by activating one of the peripheral electrodes.
A mask for producing the electrode pattern was created by cutting (using a computer numerically controlled cutter) out a stencil of the desired electrode configuration in a PET (polyethylene terephthalate) film coated with nonstick silicone. This mask was first applied over the elastomer membrane. A thin, uniform layer of SWCNTs was prepared by vacuum filtration-transfer method, in which a water dispersion of SWCNTs was passed through a Teflon filter, depositing a uniform mat of SWCNTs (Shian et al. 2012). The SWCNT mat was then transferred over the mask and membrane by pressing firmly and peeling off the filter and electrode mask, leaving behind the SWCNTs adhered to the membrane.
Figure 3-10. Metalens and dielectric elastomer actuator (DEA) mating. First fabricated separately, the (a) metalens that is transferred to an elastomer substrate is combined with (b) the DEA to form the (c) adaptive lens device.

In order to transfer the metasurface to the DEA (Figure 3-10), the metasurface and a scaffold membrane (also VHB) were first bonded together, after which the scaffold membrane was bonded to the DEA. We originally thought that the adhesion between the metasurface and membrane would be enhanced by cleaning the surface using an oxygen plasma treatment to expose dangling bonds in the form of reactive hydroxyl groups. However, we found that the opposite was true: the same dangling bonds also increased the hydrophilicity, which allowed the presence of water or humidity to infiltrate and undercut the bonding between metasurface and membrane, rendering weak adhesion. Instead, pure isopropyl alcohol (IPA) was used to clean the surface of the metasurface
Figure 3-11. Transfer process. (a) A round silicone press was used for making initial contact between the metalens on wafer to the elastomer substrate. (b) An apparatus held the silicone press and a controlled force was applied for several minutes to hours. (c) Hand pressing was also used in some cases. (d) The lens and elastomer combination was then visually inspected for any air bubbles or other defects. (e) The combination was then immersed in water so that the sacrificial layer could be released. (f) With the wafer substrate removed, the lens was then completely transferred onto the elastomer substrate.

and a chemical adhesion promoter (AP115, 3M) was used to make the metasurface slightly hydrophobic through surface silanization. AP115 was sprayed onto the metasurface and was quickly wiped off using a lint free cloth to reduce the contact time of the sacrificial layer with the small amount of water present in AP115. Although we do not know the exact chemical contents of AP115, the active ingredient is probably 3-glycidoxypropyltrimethoxysilane. Next, the membrane was pressed onto the metasurface, using a smooth, spherical press made of soft silicone. A spherical press was important to achieve good contact throughout the entire metasurface area, by overcoming the
nanoimprint proximity effect (Landis et al. 2006), by which there is less adhesion in the center than the periphery when using a flat press. The entire sample was then heated in an oven at 50 °C for 3 hours to improve adhesion by allowing the membrane polymer to flow and then cooled to room temperature.

For release, the entire sample was immersed in water (Figure 3-11). Dissolution began around the edges of the devices and moved towards the center as water slowly percolated between the nano-posts (Figure 3-12). The release time of the entire device depended on the device area and varied between minutes to hours. The scaffold membrane with the supported metasurface was then attached to the DEA by gently pressing. We refer to the scaffold membrane after combination as the intermediate
elastomer layer (IEL). Two types of layered devices were made, which we refer to as single layer (SL) and double layer (DL) devices: the SL device contains an IEL (not pre-stretched) which was cut to be as small as the metasurface, while the DL device contains an intermediate elastomer layer (pre-stretched 4x, linear) of area equal to the entire DEA, effectively increasing the stiffness of the combination.

In the final step, electrical contacts were made to the CNT electrodes. The SWCNT electrode patches were connected by conductive silver grease to leads of conductive carbon tape, which were then connected to wires (Figure 3-9).

### 3.4 Experiment and characterization

**Measurement methods**

A tunable laser (HP 8168F) operating between the wavelengths of 1440-1590 nm was used as the light source (Figure 3-13). This optical output of this laser was connected to an optical fiber collimator (Thorlabs F810APC-1550), which produced a collimated beam, 7 mm in diameter. The collimated beam was used to illuminate the device. The polarization of the beam was allowed to wander because the device was designed to be polarization independent, and no polarization dependence was observed. A high voltage source (Trek 610E) was used to tune the device. After passing through the device, the light was measured by a conventional horizontal microscope setup: a microscope objective (10x Mitutoyo M Plan Apo NIR infinity corrected objective) and a tube lens (Plano-Convex Lens, f = 200 mm) was used to magnify the beam in order to fully visualize the focal spot on the camera (digital InGaAs, Raptor OWL640). The entire horizontal microscope setup was mounted on a linear motor (NPM Acculine SLP35),
Figure 3-13. Measurement setup. Schematics (top view) of the setups for measuring the focused spot size of the device (a) with and (b) without the horizontal microscope to magnify the image. The setup without the horizontal microscope was used in at higher applied voltages, in which there was not enough space on the motorized stage to accommodate the horizontal microscope given the extended focal length of the device. (c) Setup for measuring the focusing efficiency of the device. The first aperture was used to reduce the incident beam size to 6 mm to match with the size of the device. The second aperture was used to exclude the stray light. The focusing efficiency was calculated using the ratio of the optical power at the focus (measured with both apertures and the device in place) to the total unobstructed optical power (measured with both the device and second aperture removed, but with the first aperture still in place).

which allowed horizontal scanning of the light field with a positional accuracy of 1 µm. This setup (Figure 3-13a,b) was used to characterize the focal length and focal spot size. To measure the efficiency, we replaced the microscope objective, tube lens, and camera with an optical power meter (Thorlabs PM100D). For comparison, the measured focusing
efficiency of the metasurface before transfer (fabricated on a fused silica wafer) was 91%, and the difference in device efficiency was attributed scattering and absorption by the DEA as well as imperfections in the transfer process.

Focus Measurement

Focal length (Figure 3-14, Figure 3-15) was measured by scanning a camera along the z-axis for varying voltages. From \( f_0 = 50 \text{ mm} \), the DL focal length was tuned by 15 mm (\( \Delta f / f_0 = 30\% \)) with 3 kV (Figure 3-14c) and closely followed the predicted voltage relation (Eq. 3.3), while SL displayed a greater, 107% focal length modulation (Figure 3-14c, inset). The focusing efficiency (DL), defined as the ratio of the focused optical power and the incident power, showed a high average efficiency of 62.5% with little variation throughout the tuning range (Figure 3-14b). The near constant efficiency is due to the relatively small stretching range used, which allows the posts to maintain subwavelength spacing, and to the high optical confinement factor (the calculated confinement factor, defined as \( \Gamma = P_{\text{inside, post}} / P_{\text{total}} \), is 0.71 and 0.69 ±0.01 for s=1 and 1.09, respectively) within the posts due to the large index contrast (Arbabi et al. 2015), which minimizes the effect of changes in the gap size. At 0 and 1 kV, the focal spot sizes (1/e² full beam waist) were measured and compared to the theoretical diffraction-limited spot sizes (\( \varnothing 6 \text{ mm}, M^2=1.3 \)): 34.4±1.1 \( \mu \text{m} \) (diffraction-limit: 21.4 \( \mu \text{m} \)), and 37.7±2.8 \( \mu \text{m} \))
Chapter 3: Dynamic wavefront shaping with adaptive metalenses

Figure 3-14. Measurement of tuning. (a) (Left) Z-scan of the intensity profile showing two distinct focal lengths. (Center) Image of focal spot intensity profile, i.e. x-y cross sections at the position of maximum intensity. (Right) Line scans of focal spot intensity image in the x (blue) and y (red) directions in comparison to the theoretical diffraction-limited spot size (black). (b) Additional images of the focal spots at 1.5, 2, and 2.5 kV are shown, captured directly by the camera without the microscope objective (scale bar: 200 µm). (Lower right) Measured focusing efficiency for varying voltages. (c) Measurement of focal length tuning using center electrode V5 for double layer (DL) and single layer (SL, inset) devices. Blue circles: Optical measurement of device focal length as a function of applied voltage. Solid blue line: fit of focal length data to Eq. 3.3 (R² = 0.9915). Red triangles: Measurement of stretch as a function of the applied voltage. (d) Measured Zernike coefficients of the phase profile (calculated from microscope images of lens center) showing targeted tuning of vertical astigmatism, while other Zernike coefficients exhibit little change. The large defocus value represents the designed focal power of the lens. (e) Measurement of x,y-shift control, showing two-axis displacement control from the focus position at 0 kV (yellow star) to various positions (red dots) as 1.9 kV is applied. The gray shaded area shows possible displacements, which can be accessed by concurrently activating combinations of electrodes.
(diffraction-limit: 22.7 µm), respectively (Figure 3-14a). Possible explanations for these differences include fabrication errors and small distortions introduced when the elastomer was initially pressed onto the metasurface. Due to setup constraints, images of the focal spots at higher voltages (>1 kV) were obtained directly by the camera without magnification (Figure 3-14b).

Figure 3-15. Fitting measurement of focal length tuning using electrode V5. Fitting measured focal length with applied voltage for (a) SL device and (b) DL device. Black dots: Measurement of device focal length as a function of applied voltage. Blue solid lines: fits of the relationship between the focal length and the applied voltage. Blue dashed lines: fits with 95% confidence bounds.

**Stigmator and shifter measurement**

The effect of the x stigmator, which squeezed the phase profile into an elliptical shape, was measured to show asymmetric yet spatially uniform biaxial strain according to the Fourier transform (FT) images (Figure 3-4b), indicating good stigmator performance, corroborated by the Zernike transform (Figure 3-14d). The sharpness of the FTs are
indicative of the quality of the lens, and the radii of the first order annulus, corresponding to the reciprocal space representation of the radial periodicity between posts, is a measure of the biaxial stretch. The features seen in the FTs are signatures of symmetries employed by the metasurface design algorithm.

The x,y-shifters were able to shift the image in all cardinal directions by applying appropriate control voltages (Figure 3-14e). The observed control asymmetry is likely due to residual asymmetrical stiffness around the periphery of the device, as three different structures needed to be concentrically aligned during fabrication: the outer frame, electrodes, and metasurface. Shift-induced distortions of the metasurface were minimal (Figure 3-16).
Figure 3-16. XY-shift distortion measurement. XY-shift distortion was measured by using microscope images of metasurfaces, for which displacements were produced in the up, down, left, right ((i)-(iv), respectively, with scale bar: 20 µm) directions by applying 1.9 kV to the four peripheral electrode areas. Shown below ((v)-(viii)) are the corresponding Fourier transforms of (i)-(iv), respectively. During imaging, the microscope field of view was re-adjusted with respect to the center of the metasurface after each voltage application.
Reliability test

The reliability of the device was tested with a sinusoidal voltage from 2 to 100 Hz at an amplitude of 2.5 kV. The device did not fail nor was image quality observed to degrade after >1000 cycles. The mechanical robustness of the device is attributed to the relatively small actuation strains applied. By applying a voltage square wave, the response time was measured to be 33±3 ms (Figure 3-17) and is mainly limited by the elastomer viscoelasticity, charge transfer and dissipation time in the SWCNT electrodes. Dielectric breakdown was measured at ~3.5 kV, when current begins to flow through the dielectric, damaging the device. Interestingly, the electrical breakdown was a “soft” breakdown associated with local burning through the elastomer and the same devices were able to resume operation after cycling the power. This self-healing feature is attributed to the burning and subsequent clearing of SWCNT electrodes around the breakdown location, preventing further electrical shorting.
Figure 3-17. Response time measurement. Measured rise and fall times for DL (a-d) and SL (e–f) devices. The measured data are plotted as black dots, the fits to exponential curves are shown as blue solid lines, and 95% confidence interval of the fit are shown as blue dashed lines. DL was actuated with a step voltage between 0 and 2 kV showing a (a) rise time of 33±3 ms and (b) fall time of 271±3 ms. DL was actuated with a step voltage between 0 and 3 kV showing a (c) rise time of 182±15 ms and (d) fall time of 105±13 ms. SL device actuated with a step voltage between 0 and 2 kV showing a (e) rise time of 327±7 ms and (f) fall time of 320±8 ms.
3.5 Outlook: compact adaptive optics

An electrically-tunable metalens with multiple, simultaneous control parameters, as well as a centimeter-scale metalens was demonstrated (She et al. 2017a). In comparison to existing systems, the device offers a route to extremely compact, lightweight transmissive adaptive optics with large, well-controlled tunability and high-definition (subwavelength) phase profiles. It is capable of focus tuning and aberration control simultaneously. In its current form, it is limited in speed and requires high operating voltage. Notably, this demonstration of astigmatism and image shift tuning is important for the correction of misalignment in optical systems based on metasurfaces.

By using thinner elastomers (Töpper et al. 2015) (operating voltage <12 V) or miniature high voltage components that are commonly found in devices such as mobile phones (e.g. flash modules), the voltage requirements may be substantially reduced. Speed may be further increased to the microsecond time scale (Rosset & Shea 2016). Custom-made elastomers such as silicone-based elastomers (Töpper et al. 2015) could provide a direct path for these improvements, as the elastomer dimensions and electromechanical properties could be tailored during processing. However, in addition to the consideration of mechanical stability of devices, the use of thinner or different elastomers requires adaptation and optimization of other aspects, such as the choice of stretchable electrodes and the lens transfer process.

Devices can be made using the same method with diameters ranging from microns to centimeters. These desirable characteristics are suitable for integration in size and weight-limited imaging systems and wearable devices. Our results demonstrate the feasibility of embedded autofocus, optical zoom, image stabilization, and adaptive optics;
these are expected to become essential for future chip-scale image sensors as well as head mounted optics, such as everyday eyeglasses and virtual reality hardware. Furthermore, its flat construction and inherently lateral actuation without the need for motorized parts allow for highly stackable systems, such as with stretchable electronic eye camera sensors (Ko et al. 2008), providing possibilities for new kinds of imaging systems.
Chapter 4
Dynamic polarization shaping with parallel polarization state generation

4.1 Introduction

The control of polarization, an essential property of light, is of wide scientific and technological interest. The general problem of generating arbitrary time-varying states of polarization (SOP) has always been mathematically formulated by a series of linear transformations, i.e. a product of matrices, imposing a serial architecture. Here we show a parallel architecture described by a sum of matrices.

The theory is experimentally demonstrated by modulating spatially-separated polarization components of a laser using a digital micromirror device that are subsequently beam combined. This method greatly expands the parameter space for engineering devices that control polarization. Consequently, performance characteristics, such as speed, stability, and spectral range, are entirely dictated by the technologies of optical intensity modulation, including absorption, reflection, emission, and scattering. This opens up important prospects for polarization state generation (PSG) with unique performance characteristics with applications in spectroscopic ellipsometry, spectropolarimetry, communications, imaging, and security.
Chapter 4: Dynamic polarization shaping with parallel polarization state generation

Polarization is commonly represented in two forms: the Jones vector and the Stokes parameters. The Jones vector is a complex 2-element vector, which describes completely polarized light by defining the phases and amplitudes of two orthogonal electric field components. The electrical field of a plane wave propagating in the z direction can be written as the following:

$$\vec{E}(z, t) = \begin{pmatrix} E_{0x} \cdot \cos(\kappa z - \omega t + \delta_x) \\ E_{0y} \cdot \sin(\kappa z - \omega t + \delta_y) \end{pmatrix}$$

(5.1)

The complex amplitude coefficients of equation (5.1) are known as the Jones vector: $\left(E_{0x} e^{i\delta_x}, E_{0y} e^{i\delta_y}\right)$. The Stokes parameters, which were defined by G. Stokes in 1852 to mathematically describe polarized light, including partial polarization (Stokes 1852), are extremely useful today but were historically hampered by the inability to quantify optical intensity measurements. To derive the Stokes parameters, the equation of the polarization ellipse (Collett 2005),

$$\left(\frac{E_x}{E_{0x}}\right)^2 + \left(\frac{E_y}{E_{0y}}\right)^2 - \frac{2E_x E_y}{E_{0x} E_{0y}} \cos \delta = \sin^2 \delta,$$

where $\delta = \delta_y - \delta_x$, can be rearranged into

$$\left(E_{0x}^2 + E_{0y}^2\right) - \left(E_{0x}^2 - E_{0y}^2\right) - \left(2E_{0x} E_{0y} \cos \delta\right)^2 = \left(2E_{0x} E_{0y} \sin \delta\right)^2,$$

such that the grouped terms can be written as the following Stokes parameters:

$$S_0 = E_{0x}^2 + E_{0y}^2,$$

$$S_1 = E_{0x}^2 - E_{0y}^2,$$

$$S_2 = 2E_{0x} E_{0y} \cos \delta,$$ and

$$S_2 = 2E_{0x} E_{0y} \sin \delta.$$

(5.2)
Figure 4-1. Three Stokes parameters form the axes of the Poincaré sphere.

$S_0$ is the total intensity, and $(S_1, S_2, S_3)$ is the Stokes vector describing the SOP. The Stokes parameters simplify measurement of SOP enormously by requiring only 4 intensity measurements. A triangle inequality exists, in which the total intensity $S_0^2 \geq S_1^2 + S_2^2 + S_3^2$. The ratio of the length of the Stokes vector to the total intensity is the $S_0$ degree of polarization: $DOP = \sqrt{S_1^2 + S_2^2 + S_3^2} / S_0$. The Stokes vector that is normalized to a unit vector traces all possible SOPs on a mathematical object called the Poincaré sphere.

In everyday use, SOPs are commonly met in the so-called “degenerate polarizations” as linearly and circularly polarized light but are in general elliptically polarized (Clarke 1974; Shurcliff & Ballard 1964). To describe and control the polarization of light, the projections of the electric field onto an orthogonal bases and their relative phase relation must be known and are mathematically represented by the Jones vector and Stokes Parameters (Born et al. 2000; Collett 2005).

In conventional serial architectures, the polarization of an input beam, $E_{in}$, may be transformed into any arbitrary output polarization, $E_{out}$, through a product of Jones
matrices $M_n$ corresponding to variable optical elements, each of which has a degree of freedom, $\rho_n$: $E_{\text{out}} = M_N(\rho_N) \cdots M_2(\rho_2)M_1(\rho_1)E_{\text{in}}$. Commonly found implementations of serial PSGs use optical elements that introduce suitable phase shifts or birefringence, which are represented by a product of at least two Jones matrices. These include devices such as rotating waveplates (Imai et al. 1985), Babinet-Soleil compensators (Collett 2005), Berek rotary compensators (Holmes 1964), fiber coil polarization controllers (Lefevre 1980), Faraday rotators (Okoshi et al. 1985), fiber squeezers (Ulrich 1979), polarization Michelson interferometers (Takasaki & Yoshino 1969), degree of polarization generators (Lizana et al. 2015), lithium niobate electro-optics (Kubota et al. 1980); liquid crystals (Zhuang et al. 1999); and on-chip photonic circuits (Dong et al. 2014; Miller 2013). Figures of merit that characterize the performance of these devices include temporal response, stability, mechanical fatigue, insertion loss, SOP accuracy (Okoshi 1985), and operating wavelength range.

To develop a parallel architecture, we revisit the Fresnel-Arago interference laws, which state that light beams of orthogonal polarizations cannot interfere (Fresnel et al. 1866; Collett 1971). Beams that are coherent, however, create a linear superposition to produce a new SOP. In our approach, we propose PSG by combining a limited set of prepared SOPs, which we refer to here for convenience as the “Stokes Basis Vectors” (SBVs), and are not necessarily linearly independent in the conventional sense. By modulating the intensities of a number of beams corresponding to a set of SBVs and combining them, we are able to generate any arbitrary output SOP (Figure 4-2).
4.2 Theory of parallel polarization state generation

A set of SBVs labeled by $n$ can be described as follows as Jones vectors:

$$
C_n = \begin{pmatrix} C_{0xx} \\ C_{0yy} e^{i \theta_n} \end{pmatrix} e^{i \phi_n} = \begin{pmatrix} \tilde{C}_{xx} \\ \tilde{C}_{xy} \end{pmatrix}
$$

(5.3)

where $C_{0xx}$ and $C_{0yy}$ are real coefficients, $\theta_n$ is the relative phase difference between polarization components, $\phi_n$ is the global phase, and $\tilde{C}_{xx}$ and $\tilde{C}_{xy}$ are complex amplitudes of the electric field. By linearly combining $N$ SBVs of equation (5.3) multiplied by modulation parameters, $\alpha_n$ (here real and positive scalar quantities corresponding to intensity modulations when squared), the resultant electric field can be expressed as the following:

$$
E = \alpha_1 C_1 + \alpha_2 C_2 + \cdots + \alpha_n C_n
$$

(5.4)

While the global phase of each SBV, $\phi_n$, does not affect its SOP, relative phase is an important factor in the interference between the SBVs, and its physical origin is the phase shift measured at the location where beams combine; $\phi_n$ can be tuned by changes in optical path length or by other means, such as resonant optical elements. It is shown later that the combination of a minimum of four SBVs is required to generate arbitrary SOPs, so that any desired Stokes vector can be mapped to four modulation parameters: $(S_1, S_2, S_3) \rightarrow (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.
Figure 4-2. Concept of parallel polarization state generation. (a) An illustration showing the general, modular implementation of the described method for a parallel polarization state generator (PSG). An input beam is (i) split into four beams of different polarizations, which are then (ii) intensity modulated either in reflection or transmission, (iii) and finally combined to form a single output beam, the polarization and phase of which can be tuned with a precision and speed limited by the modulator. (b) A schematic of PSG architecture is shown, in which modulators are placed after light sources Ai with well-defined states of polarization (SOP) and relative phase, and their weighted linear superposition produces the desired output signal. (c) Generation of horizontally polarized light using this method is illustrated. The electric fields of four propagating electromagnetic waves (red, green, blue, and yellow) with elliptical polarizations are superimposed and plotted as function of wave propagation position. They are intensity modulated and beam combined to generate the desired horizontal polarization (black).
Figure 4-3. Simulations of polarization state coverage. Two systems with distinct sets of Stokes basis vectors (SBVs) were simulated— one composed of degenerate SOPs and the other with SOPs mapped to a regular tetrahedron on the Poincaré sphere. Intensity modulation parameters of individual SBVs are varied to generate polarization trajectories and whole regions of accessible SOPs. (a) A system with SBVs with four degenerate SOPs: linear horizontal (C1), vertical (C2), +45° with a 180° phase shift (C3), and right circular polarization (C4), are shown. A Monte Carlo simulation (blue points) was performed by randomly varying the intensity modulation parameters and showed complete, yet non-uniform coverage of SOPs over the Poincaré sphere. A polarization trajectory between SBVs C3 to C4 is shown for coherent combination (blue line) and incoherent combination (red line). Incoherent trajectories are geodesics. (b) A system with SBVs optimized for better uniformity of SOP coverage is shown, in which the SBVs are vertices of a regular tetrahedron inscribed in the Poincaré sphere, as opposed to the degenerate SBVs. In Jones vector notation, the SBVs used here were [0.7071, 0.7071i], [-9.856, 0.1691i], [0.5141, 0.7941-0.3242i], and [0.5141, -0.7941-0.3242i], labeled as C1-4, respectively.
In the case of four SBVs, equation (5.4) can be rewritten as the following real matrix equation:

\[ \begin{pmatrix}
C_{1x}^{\text{re}} & C_{2x}^{\text{re}} & C_{3x}^{\text{re}} & C_{4x}^{\text{re}} \\
C_{1x}^{\text{im}} & C_{2x}^{\text{im}} & C_{3x}^{\text{im}} & C_{4x}^{\text{im}} \\
C_{1y}^{\text{re}} & C_{2y}^{\text{re}} & C_{3y}^{\text{re}} & C_{4y}^{\text{re}} \\
C_{1y}^{\text{im}} & C_{2y}^{\text{im}} & C_{3y}^{\text{im}} & C_{4y}^{\text{im}}
\end{pmatrix}
\begin{pmatrix}
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\alpha_4
\end{pmatrix} =
\begin{pmatrix}
E_x \cos \phi \\
E_x \sin \phi \\
E_y \cos (\theta + \phi) \\
E_y \sin (\theta + \phi)
\end{pmatrix}
\]

(5.5)

where $\theta$ and $\phi$ are defined as in equation (5.3). This can be solved for real and positive $\alpha_n$ given a set of SBVs represented by the square matrix on the left hand side and the desired SOP given by the right hand side. The square values of the calculated $\alpha_n$ are used to modulate the intensities of the SBVs for final PSG. Additionally, the number of SBVs can be increased and each prepared with well-defined $\phi_n$ in order to add the capability of phase control to the generated SOP.

Polarization modulation can be visualized as dynamic polarization trajectories on the surface of the Poincaré sphere (Figure 4-3a,b). For example, the linear combination of any two SOPs can be varied in order to create a line of SOPs on the Poincaré sphere: $E = \alpha C_1 + (1-\alpha)C_2$, in which two SOPs, $C_1$ and $C_2$ (that could be SBVs), are parameterized by $\alpha$ that is varied from 1 to 0 (Figure 4-3a). Combining SOPs generates new SOPs by way of interference; depending on their relative phase, paths with varying curvature can be generated (Figure 4-4). In order to deviate from this path, a third SOP, $C_3$, must be introduced to provide one more degree of freedom, which expands the generable SOPs from a line to a surface (region).
Chapter 4: Dynamic polarization shaping with parallel polarization state generation

Figure 4-4. Phase dependence of coherent polarization trajectories on the Poincaré sphere. The effect on polarization trajectories by changing the relative phase between two linearly combined SOPs is shown. The polarization trajectory connecting SOPs can be modified in either direction by controlling the relative phase.

Within an arbitrary set of SBVs, each subset of three SBVs (C1, C2, and C3) can generate a surface bounded by the trajectories connecting each pair of SBVs (C1 and C2, C1 and C3, C2 and C3). Then arbitrary trajectories can be generated within this allowable surface, such as spiral or even chaotic trajectories (Figure 4-3c,d and Supplementary Information). In the case of coherent combination, we obtain a trajectory that is sensitive to the relative phase between SBVs (Figure 4-3c). In contrast, the combination of SOPs with greatly reduced mutual coherence, i.e. incoherent, traces a trajectory corresponding to the shortest path (geodesic) connecting the SOP of the initial to the final state on the Poincaré sphere, which is independent of relative phase.

Coverage of the entire Poincaré sphere by SBVs comprised of four degenerate SOPs (the horizontal, vertical, +45°, and right circular polarizations) is shown in Figure 4-3a,c. The regions enabled by each subset of three SBVs piece together to entirely cover the Poincaré sphere. The degenerate SBVs take on any four of the six degenerate polarizations, which are the linear horizontal, linear vertical, linear +45°, linear -45°,
Chapter 4: Dynamic polarization shaping with parallel polarization state generation

Figure 4-5. Regions of coverage spanned by Stokes Basis Vector subsets. (c) The degenerate SBV system is mapped using a Mercator projection of the Poincaré sphere, where $\theta$ is the polar angle and $\phi$ is the azimuthal angle. All coherent and incoherent polarization trajectories between SBVs are shown with black dotted and red solid lines, respectively. The coherent polarization trajectories connected to C1 are warped by increasing the relative phase difference between C1 and other SBVs by $6^\circ$ (blue dotted lines). The colored regions show the regions of SOPs enabled by combinations of three SBVs: by combining C1, C2, and C4, with varying intensity modulation parameters, all SOPs in the blue region can be generated; similarly, combinations of (C1, C3, C4) and (C2, C3, C4) generate the red and green regions, respectively. However, (C1, C2, C3) generate a region of no area because these SBVs are not linearly independent in this particular system. (b) The regular tetrahedron SBV system. Coherent and incoherent polarization trajectories between SBVs are shown with black and red dotted lines, respectively. In this case, regions of SOPs generated by combinations of sets of three SBVs are well distributed and have similar size and great overlap, yielding better uniformity. The overlapping regions are color coded and labeled as the following: C1, C2, C3 combine to cover regions a, b, and c; similarly: C1, C2, C4 (a, d, e); C1, C3, C4 (c, e, f); and C2, C3, C4 (b, d, f).
right circular, and left circulation polarizations. We explored a system of SBVs in the four following polarizations: linear horizontal, linear vertical, linear +45°, and right circular. Coverage of the Poincaré sphere by possible SOP states using the above system is shown in Figure 4-3a,c. Figure 4-6 plots the Mercator projection of Figure 4-3c on the Poincaré sphere. Each system is uniquely defined by the SOPs and the global phases $\phi_n$ of the SBVs that comprise it. In this particular system, all SBVs have $\phi=0^\circ$, except the SBV with linear +45° SOP, for which $\phi=180^\circ$.

Figure 4-6. Degenerate SBV coverage. The Poincaré sphere is shown, covered by possible SOPs, as generated by linear combinations of four degenerate SBVs, in the following polarizations: linear horizontal, linear vertical, linear +45°, and right circular. All SBVs have global phases $\phi=0^\circ$, except that of the linear +45° polarization with $\phi=180^\circ$. 
However, SOP coverage (the angular change in SOP corresponding to a change in modulation parameters) is nonuniform for the set of degenerate SBVs. We improved uniformity by borrowing from optimization techniques used in polarimetry (Azzam et al. 1988; Sabatke et al. 2000; Renes et al. 2004): optimal and minimal polarimetry and symmetric informationally complete positive operator valued measures (SIC-POVM). In these methods, a polarimeter measures the intensities of four states corresponding to the vertices of a regular tetrahedron inscribed in the Poincaré sphere. This arrangement maximizes the distance between measured states. When constructing a PSG with degenerate SBVs, the four SOPs define an irregular tetrahedron, resulting in a greater density of SOPs gathered around octant I of the Poincaré sphere. We calculated that a set of SBVs with elliptical SOPs defining a regular tetrahedron greatly improves uniformity of coverage compared with four degenerate SBVs (Figure 4-3b,d).

The uniformity of the system can be described by the metric on the Poincaré sphere, where angular separation between SOPs in Stokes space is given by $\cos \theta_{AB} = \frac{S_A \cdot S_B}{||S_A|| ||S_B||}$ and in Jones vector space $\cos^2 \theta_{EiEj} = \frac{\langle E_i|E_j \rangle^2}{\langle E_i|E_i \rangle \langle E_j|E_j \rangle}$. By describing nearby states as $E_i = C_{ij} \alpha_j$ and $E_i = C_{ij} (\alpha_j + \delta \alpha_j)$, it is possible to construct a metric tensor, where $ds^2 = d\theta^2$. For the case of coherent combination, this is the Fubini-Study metric, and for incoherent combination this is the Bures metric.

As described in the main text, coherent combination (Figure 4-7) produces polarization trajectories that are sensitive to the difference in global phases $\phi_n$ between SBVs (Figure 4-4), whereas incoherent combination necessitates a geodesic trajectory that is insensitive to $\phi_n$ (Figure 4-8). It is possible to switch between these two...
combination methods to generate trajectories with degrees of coherence intermediate between the two limits by changing the mutual coherence between SBVs. Mutual coherence, or the degree of cross-correlation, can be tuned by varying the optical path length between SBVs. By having the optical length exceed the coherence length of the light source, the SBVs no longer have a fixed phase relation and incoherently combine. Figure 4-8 further illustrates the difference in polarization trajectories as generated by the two methods of combination.

![Figure 4-7. Coherent polarization trajectories on the Poincaré sphere. Trajectories are shown, where the polarization is varied from the initial state (red) to the final state (blue). Four trajectories between degenerate SOPs are plotted: a) linear horizontal to linear vertical, b) right circular to linear horizontal, c) linear vertical to linear +45°, and d) linear +45° to right circular. The global phase of the linear +45° polarization is defined to be 180° out-of-phase with respect to the other degenerate SOPs, such that in Jones vector notation it is [-1, -1].](image-url)
The SOP measured following incoherent combination can be described as a linear sum of Stokes vectors:

\[ S_{\text{out}} = \alpha_1^2 S(C_1) + \alpha_2^2 S(C_2) + \alpha_3^2 S(C_3) + \alpha_4^2 S(C_4) , \]

in which \( S(C_i) \) is the Stokes vector corresponding to the SOP of \( C_i \), and so on. This can be seen as the simultaneous detection of the intensities of non-interfering beams with their unaltered SOPs; hence the intensities of Stokes vectors add linearly on the detector side. The incoherent trajectory is insensitive to the relative phase difference between SBVs. Finally, the degree of polarization of the generated SOP as measured by a polarimeter has two contributors: (a) any unpolarized background originating from the unpolarized parts of the sources’ signals and (b) any less than unity value of the degree of cross-correlation (mutual coherence) between combined SBVs.
Figure 4-8. Coherent and incoherent polarization trajectories on the Poincaré sphere. a) Various perspective views of an example system. The SBVs used are labeled C1-4. Polarization trajectories are generated by coherent combination (blue line) and incoherent combination. b) This example system is implemented using the experimental setup described in the main text. A polarization trajectory is generated by coherent combination by keeping the optical path length between the two SBVs C2 and C3 well below the coherence length of the laser (~20 cm). The trajectory is measured by the polarimeter and the data are shown. c) An incoherent polarization trajectory following the geodesic path between SBVs C2 and C4 is generated by making the optical path length between SBVs much longer than the coherence length of the laser (hence reducing the mutual coherence).
4.3 Device design and assembly

To implement our method experimentally, a wide range of intensity modulators and wavelengths, as well as free-space, guided, and on-chip configurations are available to us. In our experiment, we used a digital micromirror device (DMD) to modulate four spatially separated SBVs derived from a laser beam to digitally generate a laser beam with arbitrary SOP (Figure 4-9). In order to modulate the intensities of each of the four beams, a black and white image corresponding to a random binary matrix with an average value equal to the desired intensity modulation parameter was displayed on each quadrant of the DMD. The DMD was a Texas Instruments DLP3000. The displayed image was changed according to the desired SOP. The output was then measured using a free-space polarimeter (Thorlabs PAX5710).
Figure 4-9. Experimental setup. Laser light is prepared in the linear +45° polarization using a wire-grid polarizer. The beam is split into two beams by a non-polarizing beam splitter (BS). Each of these beams is split again using variable circular polarizers (VCPs) into elliptical polarization states, which can be tuned by rotating the internal quarter wave plates. Variable neutral-density filters (VNDFs) are placed directly after the VCPs to balance the four beam intensities. The four beams are then directed onto four quadrants of the surface of a computer controlled Texas Instruments DLP3000 digital micromirror device (DMD). The DMD is composed of an array of polarization-insensitive mirrors that can be switched in one of two positions. Mirrors that point in the direction of the output beam contribute to the total intensity and all other light is directed into a beam dump. The DMD behaves as a 2-D diffraction grating for the incident laser light. An iris is used to select the strongest diffraction order. The path length differences of the four intensity-modulated beams passing through the iris are adjusted to be less than the coherence length of the laser (< 20 cm) with a series of mirrors. They are combined using three non-polarizing beam splitters to form a single beam. Finally, this beam is passed through a 100-µm pinhole, in order to select a small uniform portion of the wavefront of the combined beam to maximize the degree of polarization, to form the PSG output.
4.4 Experiment

In experiment, coherent trajectories were generated between SBVs (Figure 4-10a). A Monte Carlo experiment was performed to probe coverage of SOPs over the Poincaré sphere with 200 random modulation parameters and produced good uniformity of coverage using a set of regular tetrahedral SBVs (Figure 4-10b). A time-varying polarization signal was measured at slow speeds and matched well with the theory based on equation (5.5) (Figure 4-10c). Measurements were also performed of the switching speed between linear horizontal and vertical SOPs, in which a high-speed pseudorandom bitstream was displayed on the DLP chip to generate an eye pattern (Figure 4-10d).

Sources of error include vibration of optical components. The final polarization state is sensitive to the jitter in the relative phase between each of the four beams, and the average angular SOP error was measured to be 5.9° on the Poincaré sphere (Figure 4-10a,c). The SOP profile along the interfering wavefront changes smoothly, due to slight misalignment between the four beams, causing the relative phase difference between the SBVs to vary slightly as a function of position. Vibration of the pinhole causes the output beam to be a sample of a changing portion of the preceding wavefront and leads to SOP error. Additionally, simultaneous sampling of multiple SOPs by the pinhole leads to multiple SOPs detected and integrated by the polarimeter, which decreases the degree of polarization, as can be seen with unpolarized light that is mathematically decomposed into two uncorrelated orthogonal elliptical SOPs (Collett 2005).
Figure 4-10. Experimental results. (a) The Stokes basis vectors (SBVs) are set to SOPs approximating a regular tetrahedron on the Poincaré sphere. The SBVs C1-4 were measured and the resulting tetrahedron is drawn. Coherent polarization trajectories between all SBVs are generated by modulating intensities in 20 discrete increments spanning 20 seconds, and the data as measured by polarimeter are shown. (b) The results of a Monte Carlo experiment, in which 200 random intensity modulation parameters $\alpha$ were used, are shown on the Poincaré sphere, indicating good uniformity of coverage of SOPs. (c) Time series data of a coherent polarization trajectory between two SBVs (C2 to C4) in (a) are compared to theoretical calculation (dotted line) and show good agreement. S1, S2, and S3 are elements of the Stokes vector. (d) An eye pattern is generated for a polarization signal that switches between linear horizontal and vertical polarizations using the DLP3000. The data are shown for a pseudorandom bitstream modulated at 1 kHz. The inset shows the measured settling time (eye rise and fall time) to be 3.5 $\mu$s following an exponential.
The polarization-modulated beam was incident on a high-speed photodiode (Thorlabs DET100A) with a mounted linear polarizer, and the optical signal was measured on an oscilloscope (Agilent 54855A DSO) triggered by the automatic trigger signal of the DLP controller. Switching speed was measured up to the maximum speed allowed by the DLP3000 at 4 kHz without any degradation or impact on SOP signal quality. The measured settling time was extremely fast (3.5 µs), following an exponential for a 1 kHz bit stream, which reflects the settling time of the DMD. SOP noise was dominated by the instability of relative phase between interfering beams, which are best seen in the polarization trajectory measurements of Figure 4-10a,c.

4.5 Outlook: challenges and applications

Our main concern with the parallel architecture, yet, is insertion loss. Absorption or reflection modulators inherently use loss as a means of modulation. Additionally, coherent beam combining methods can only efficiently combine beams that are in-phase and have equal amplitude (Fan 2009), and our architecture rarely combines beams that satisfy both requirements. However, improvements can be made easily to the modulation stage by using directional couplers (Lu et al. 2015) that retain all of the optical power when setting the relative modulation parameters between the SBVs. In the combination stage, a more sophisticated method is still sought to combine beams of varying amplitudes. Nonetheless, numerical calculations show that loss due to coherent beam
Figure 4-11. Insertion loss calculation. Insertion loss is calculated for 80,000 SOPs distributed uniformly over the Poincaré sphere. The distribution of insertion losses are represented by histograms, for which a set of 4 degenerate SBVs and a set of regular tetrahedral SBVs are shown in blue and orange, respectively.

combining is at a level that may be acceptable for applications in which the features of parallel polarization state generation are desirable. The average theoretical insertion loss by generating 80,000 SOPs distributed uniformly over the Poincaré sphere was calculated to be $6.5 \pm 4.4$ dB for a set of 4 degenerate SBVs and $8.0 \pm 2.1$ dB for a set of regular tetrahedral SBVs.

The fundamental limitation to the efficiency of the parallel architecture stems from inefficient beam combining. The theoretical coherent beam combining efficiency for a system with $N$ ports has been derived as the following (Fan 2009):
where $P_m$ is the power and $\phi_m$ the phase of the $m$th beam. This assumes perfect alignment and coherence, so, in practice, there will be additional losses. Theoretical insertion loss ($IL$) was calculated using equation (5.6), where $IL = -10\log(\eta)$, for a large number of SOPs which represented uniform coverage of the Poincaré sphere. Descriptive statistics are shown in Table 4-1.

\[
\eta = \frac{1}{N} \sum_{m=1}^{N} \sqrt{P_m \exp(j\phi_m)}^2
\]

\[(5.6)\]

<table>
<thead>
<tr>
<th></th>
<th>Degenerate SBVs</th>
<th>Regular tetrahedral SBVs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>0.22</td>
<td>0.16</td>
</tr>
<tr>
<td>(linear)</td>
<td>6.50</td>
<td>8.04</td>
</tr>
<tr>
<td>Insertion loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(dB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Minimum</td>
<td>13.05</td>
<td>11.89</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.65</td>
<td>0.28</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>4.47</td>
<td>2.11</td>
</tr>
</tbody>
</table>

Table 4-1. Insertion loss descriptive statistics accompanying Figure 4-11.
A PSG constructed from a set of 4 degenerate SBVs has a greater range of insertion loss compared to that of regular tetrahedral SBVs (Figure 4-11 and see standard deviation values in Table 4-1). This can be explained by the increased uniformity of the regular tetrahedral SBVs, which are well-separated SOPs in state space, in contrast to the degenerate SBVs, implying that on average each gengerable SOP will be much closer to one single SBV than the other three, leading to a large yet frequent power imbalance between the beams when combining. With the set of 4 degenerate SBVs, there is greater variability in the state space distance from each gengerable SOP to the SBVs.

Performance characteristics of our implementation are promising, with an SOP settling time (representing speed and stability) of 3.5 µs compared to a state-of-the-art device (Thorlabs DPC5500) with 150 µs for < 10° deviation and 1 ms for < 1° deviation. However the SOP accuracy of our embodiment (5.9° error) is limited by the unstable relative phase between the four SBVs, whereas the DPC5500 can be as accurate as 0.25°. However, there is room for major improvements, in terms of both speed and accuracy, such as by using faster modulators and miniaturization; the latter would greatly increase the phase stability between SBVs and reduce the error. Realistically, we expect PSGs in the visible and telecom wavelengths, for example, to achieve the speeds of the fastest modulators available, e.g. greater than 40 GHz (lithium niobate), pushing PSG technology from the kiloradians/second regime into the gigaradians/second.

In conclusion, we have introduced and experimentally implemented a parallel architecture for PSG, based on intensity modulation of separate polarization components. A major advantage is that the particular features of an embodiment are determined by the technology of intensity modulation used. For example, in our case, broadband metallic
mirrors of the DMD used would translate to broadband PSG. By using other modulator technologies or materials, the method can be adapted for the desired wavelength. Furthermore, figures of merit, such as speed and affordability, will continue to increase commensurately with modulator development: e.g., a system built with injection-locked directly modulated lasers (Liu et al. 2014). It is interesting to note that the architecture can be inverted to form a conventional Stokes polarimeter, suggesting a polarization transceiver. In addition to foreseeing new applications in science and technology, analogous interference phenomena exist in quantum mechanics (as can be seen by the mathematical relationship of the Pauli matricies (Fano 1954) and the coherency matrix (Born et al. 2000) with the Stokes parameters, as well as the Bloch sphere with the Poincaré sphere), which may provide the potential to generalize this method to two-level quantum systems, such as coherent electronic and magnetic systems.
Bibliography


Clarke, D., 1974. Nomenclature of Polarized Light: Linear Polarization. Applied Optics,


Lu, Z. et al., 2015. Broadband silicon photonic directional coupler using asymmetric-


Olles, J.D. et al., 2011. Optical performance of an oscillating, pinned-contact double


