Assume Your Neighbor is Your Equal:
Inverse Design in Nanophotonics

A dissertation presented
by
Raphaël Jean-Marie Fernand Pestourie
to
Harvard 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 Mathematics

Harvard University
Cambridge, Massachusetts
December 2019
Assume Your Neighbor is Your Equal: Inverse Design in Nanophotonics

Abstract

Metasurfaces are optical devices designed for a range of wavelengths. They are subwavelength patterned surfaces with dimensions much larger than the wavelength. These two characteristic length scales are orders of magnitude apart, which makes simulations and optimization of metasurfaces a costly supercomputer scale problem with state-of-the-art solvers. Previously, the community has circumvented this problem by making a local periodic approximation and has simplified the problem into a phase matching problem for fully-known solutions of the Maxwell equations. In this dissertation, I reframe the metasurface design problem as an optimization problem. Based on the local periodic approximation—which I studied analytically, I derived an approximate hybrid solver for metasurfaces which solves orders of magnitude faster than the state of the art. This approximate solver relies on a surrogate model for the solution of each pattern. To find the best surrogate model, I frame the problem as a data-driven artificial intelligence problem. In this research, I designed tangible devices: a lens with extended depth of field and lenses correcting for chromatic aberration in the visible range, with diameters of up to 1 cm. Resulting metasurfaces were fabricated and measured, successfully validating the design framework in both two and three dimensions. This dissertation is a journey at the interface of applied mathematics, applied physics, and artificial intelligence.
Contents

1 Introduction 1

2 Metalens optimization 7
  2.1 Local Periodic Approximation 8
  2.2 Existing methods of design 10
  2.3 “Assume your neighbor is your equal: inverse design in nanophotonic” 14

3 Inverse design of large-area metasurfaces 17
  3.1 Introduction and motivation 17
  3.2 Locally periodic approximation 19
  3.3 Metasurface inverse design methods 25
  3.4 Applications: RGB lens, demultiplexer, and angle-insensitive lens 30
  3.5 Beyond subwavelength periods 35
  3.6 Concluding remarks 36

4 Lens with extended depth of field 40
List of Tables

2.1 Benchmark for inverse design in nanophotonics: the first row corresponds to the analytic methods, the second to inverse design with brute force solver, the third to our optimization framework. The first column corresponds to gradient based optimization, the second to the ability to simulate a metasurface, the third to the ability to design for objective that do not have an analytic solution, the fourth to the capacity to handle multiple constraints optimization. Our approach (Inv. Design/ LPA) fills the gap for inverse design of metasurfaces compared to the state of the art.  

4.1 Comparison between the ordinary and EDOF metalenses for 66.66μm, 100μm, 133.33μm metalenses, respectively.  

5.1 Minimum NA needed before a diffractive order contributes to the focal spot.  

5.2 Specifications of the achromatic lens for red, green, and blue: [488, 532, 658] nm with 2 mm diameter.  

5.3 Specifications of the achromatic lens for six wavelength across visible range with 2 mm diameter.
5.4 Specifications of the achromatic lens for red, green, and blue with 1 cm diameter. 69

6.1 Support of parameters in the H-shaped unit cell 75
Listing of figures

2.1  High-order corrections to the ray optics approximation 2.1. (a) Geometrical configuration of the problem under consideration which corresponds to an unit amplitude plane-wave impinging at normal incidence on a metasurface that renders a transmitted plane-wave with wave-vector forming an angle of 22.5° with respect to the metasurface. (b) and (c): Real part of the total field 0–th and 1st order approximations. (d), (e), (f), (g) and (h): Absolute errors $|u - u_{N}^{\text{tot,ff}}|$, for $N = 0, 1, 2, 3$ and $4$, respectively, in the zeroth (no correction), first, second, third and fourth order corrections to the ray optics approximation $u_{0}^{\text{tot,ff}}$. The color scales were adjusted according to the maximum error displayed in each one of the figures.  

2.2  Total field solution of the problem of scattering of a Gaussian beam by a metasurface with interface impedance function engineered to allow for surface wave modes to propagate along the metasurface.
2.3 Monochromatic 3D metalens (NA = 0.37). A few portions of the lens have been magnified for easy viewing; note the scale bars. The lens consists of a single TiO$_2$ layer above the silica substrate. The far field profile is obtained by locally periodic approximation, showing diffraction-limited focusing.

3.1 Schematic of our design method: exact Maxwell scattering solutions for a set of periodic unit cells (top) are composed into an approximate solution for an arbitrary aperiodic composition (right), and this approximation is then used for large-scale optimization to determine the metasurface parameters to maximize a given objective (e.g. the focal intensity, bottom). The performance of the final design can then be fed back into adjusting the design of the unit cell (left).

3.2 Left: an arbitrary aperiodic metasurface (top) is approximated by solving a set of periodic scattering problems (bottom), one for each unit cell, to obtain the scattered field just above the surface (horizontal line segments). Right: 0th diffracted-order amplitude (top) and phase (bottom) of periodic subproblems as a function of the pillar width. This is precomputed for several widths (markers) and interpolated as needed.

3.3 Bottom: geometry of a metasurface designed for a 5-degree incident plane wave of wavelength 532 nm and focal length 14.7 μm (numerical aperture of 0.3) using the wavefront method. This design produces a field with the needed phase (middle). Top: $|E_z|^2$ intensity plot shows focusing to the target focal spot.
3.4 Left: $|E_2|^2$ intensity plot of the scattered field of a metalens using our locally periodic approximate solver showing good agreement with a brute force calculation (middle). Right: the field sections computed by the two solvers show perfect agreement close to the focal lines (top). This agreement starts to deteriorate the closer the section is to the metasurface (bottom).

3.5 Bottom: the focal line of the scattered field for the three target wavelengths (blue, green, and red) show a clear focusing on the target focal axis. Middle: sensitivity plot for the focal length with respect to the wavelength show chromatic aberration, and each wavelength objective creates a “spurious focus” (local maximum along along the focal axis) on the focal axis at other wavelengths. The red spots represent the foci for each wavelength, and we clearly see chromatic aberration. Top: intensity at the target spot vs. wavelength.

3.6 Bottom: focal lines for the three target wavelengths (blue, green and red) focus on points sixty microns apart. Top: the field produced by our design focuses on the desired foci, the high-intensity regions for blue (left) and red (right) are tilted because their foci are off-axis.

3.7 Left: the focal lines for 0-degrees, 3-degrees, 6-degrees, and 9-degrees angles of incidence show four foci with the maximum intensity at the target focal spot (at $x = 0$), the other three peaks correspond to the other three target angles. Right: corresponding produced field around the foci, the focal spot becomes more tilted as the angle of incidence increases.
3.8 Bottom: the geometry (left) from intensity optimization shows big variations in the width of the pillar, and produce good focusing when simulated with a brute force simulation (right), or our locally periodic solver (middle) which includes the diffractive orders $\pm 1$. Top: the geometry (left) from wavefront optimization shows poor focusing both using our locally periodic solver (middle) or a brute force calculation (right). All the intensity plots have the same color scale.

4.1 Scanning electron micrograph (SEM) of (top-left) inverse-designed EDOF metalens, (top-right) traditional metalens; (middle-left and middle-right) Zoom-in SEM on inverse-designed EDOF metalens which shows silicon nitride gratings forming cylindrical EDOF metalens; The scale bars correspond to 10 $\mu$m (top-left and top-right) and 1 $\mu$m (middle-left and middle-right), (bottom) Con-focal microscopy setup used to measure the metalenses.

4.2 Simulated intensity along the optical axis of the ordinary metalens for 66.66 $\mu$m (A), 100 $\mu$m (B), and 133.33 $\mu$m (C) focal lengths. Experimentally measured filed profile along optical axis for EDOF metalenses with 66.66 $\mu$m (D), 100 $\mu$m (E), and 133.33 $\mu$m (F) focal lengths. (G-O) show cross-section of the beam size in different distance from EDOF metalenses with their Gaussian and double Gaussian fit functions. All numbers which are measured experimentally has $\pm 2$ $\mu$m error bar.

4.3 Performance of fabricated ordinary and EDOF metalenses as a function of distance along the optical axis. Measured FWHM of ordinary and EDOF metalenses as a function of distance corresponds to 66.66 $\mu$m (A), 100 $\mu$m (B), and 133.33 $\mu$m (C) metalenses. Measured focusing efficiency of 66.66 $\mu$m (D), 100 $\mu$m (E), and 133.33 $\mu$m (F) metalenses as a function of distance along the optical axis.
5.1 (A) 2 mm diameter lens unit cell with TiO$_2$ nanofin on top of a silica substrate with period 400 nm. Each unit cell has two parameters—the lengths and a width of the nanofin—and is fabricated using e-beam lithography. (B) 1 cm diameter lens unit cell with silica nanopost on top of a silica substrate with period 850 nm. Each unit cell has one parameter—the radius of the post—and is fabricated using UV lithography.

5.2 (A) parameters function for a 2 mm diameter lens, (B) parameters function for a 1 cm diameter lens both achromatic for red, green, and blue.Insets show zooms of the parameter function at the center and for larger radius, in both cases the parameter is oscillatory while remaining non-trivial.

5.3 Setup for measurement the SuperK continuum laser is collimated on the metalens. The scattered field is then collected by a objective lens, which enables measurement of the focus of a high NA lens. The light is recollimated and then focused on the camera. The measuring equipment is set on a translational piezoelectric stage to enable precise control of its position.

5.4 We validated our approximate solver against brute force simulation with the 2 mm unit cell on a single wavelength design for 530 nm and a 10 $\mu$m diameter. The FDTD simulation (A-B) shows clear focusing both on the $xz$-cut (A) and along the focal axis (B). Because of symmetry, the approximate solver is only fit to compute the intensity along the focal axis (C). The two solvers match very well in the far field, and as expected, differ significantly close to the metasurface (at a distance less than 5 $\mu$m away from the metasurface). (D) We optimized for a lens with 2 mm diameter (with a refinement of 3 per wavelength) and for red, green, and blue. The intensity along the focal axis shows clear focusing at the focal spot for the three desired wavelengths [488, 532, and 558] nm. The numerical aperture of this lens is 0.2 (focal length of 2.5 cm with diameter of 1 cm).
5.5 Measurement results of focusing intensity distribution of the metalens in the XZ cross section at three designed wavelengths: 488 nm, 532 nm and 658 nm under different polarization configurations. (x-direction is along lens radius and z-direction is along the optical axis). The incident beam is coupled from a SuperK continuum fiber laser and has a bandwidth of \(~\)5 nm. (Left) The incident beam is unpolarized and focused by the metalens. The output beam intensity profile is captured by the CCD camera through an optical relay system without passing through any polarizer. (Middle) The incident beam is selected to be left-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the right-handed circularly-polarized component. (Right) The incident beam is selected to be right-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the left-handed circularly-polarized component.

5.6 (A) We characterize the lens by measuring the intensity of the field along a line parallel to the metasurface and going through the focal point. Then we fit an Airy function to compute the Strehl ratio. A Strehl ratio bigger than 0.8 signifies that the lens is diffraction-limited, which it is for the three wavelengths [488, 532, and 658] nm when converting from RCP to LCP. (B) We show the intensity of the field measured on a cut parallel to the metasurface and going through the focal point. We see an error pattern compared to an Airy disk with a \(C_4\) symmetry. It comes from the squared lattice which breaks the \(C_\infty\) symmetry of the parameter function. The error from the local periodic approximation has the same symmetry as the lattice.
5.7 (A) We characterize the lens by measuring the intensity of the field along a line parallel to the metasurface and going through the focal point. Then we fit an Airy function to compute the Strehl ratio. A Strehl ratio bigger than 0.8 signifies that the lens is diffraction limited, which it is for the three wavelengths [488, 532, and 658] nm for unpolarized light. (B) We show the intensity of the field measured on a cut parallel to the metasurface and going through the focal point. We see an error pattern compared to an Airy disk with a $C_4^v$ symmetry. It comes from the squared lattice which breaks the $C_{\infty}^v$ symmetry of the parameter function. The error from the local periodic approximation has the same symmetry as the lattice. Apart from 658 nm wavelength which shows more side lobes, the results are very similar to the conversion from LCP to RCP.

5.8 Optical microscope images of the fabricated metalens. Top panel: optical microscope images at 50× magnification. Bottom panel: optical microscope images at 100× magnification. As can be seen from the images that the fabricated metalens suffers from minor stitching errors. The stitching errors occur at the boundaries of e-beam writing fields (area of 250 μm × 250 μm) during e-beam lithography process. The stitching error line is across the outer part of the metalens and will introduce wavefront errors. The rest of the metalens is intact. We annotated the stitching error boundaries with red arrows.

5.9 The intensity along the focal axis show clear focusing for our designed lenses achromatic for [490, 520, 540, 570, 610, and 650] nm. (A) The numerical aperture of this lens is 0.1, so a diameter of 2 mm corresponds to a focal length of about 1 cm, the focal axis shows clear focusing, but also presents spurious foci. (B) The numerical aperture of this lens is 0.2, so a diameter of 2 mm corresponds to a focal length of about 5 mm. This lens shows similar features as (A).
5.10 Measurement results of focusing intensity distribution of the metalens of $\text{NA} = 0.2$ in the $xz$-cross section at six designed wavelengths: 490 nm, 520 nm, 540 nm, 570 nm, 610 nm, and 650 nm under different polarization configuration. ($x$-direction is along lens radius and $z$-direction is along the optical axis). The incident beam is coupled from a supercontinuum (SuperK) fiber laser and have a bandwidth of $\sim 5$ nm. (a) The incident beam is unpolarized and focused by the metalens. The output beam intensity profile is captured by the CMOS camera through an optical relay system without passing through any polarizer. (b) The incident beam is selected to be left-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the right-handed circularly-polarized component. (c) The incident beam is selected to be right-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the left-handed circularly-polarized component.

5.11 Measured intensity profiles at the focal planes of the 6-wavelength achromatic metalens for design wavelengths of 490 nm, 520 nm, 540 nm, 570 nm, 610 nm, and 660 nm, respectively. The incident light is from a supercontinuum (SuperK) laser and is unpolarized. The bandwidth of the light source is around 5 nm. The intensity profiles of the focal spots are captured by a CMOS camera through an optical relay system (i.e. microscope setup) without passing through any polarizer.
5.12 Measured intensity profile of the metalens in line cut through the center of focal spots at design wavelengths of 490 nm, 520 nm, 540 nm, 570 nm, 610 nm, and 650 nm. The solid lines are measurement results and the dashed lines are intensity profiles of ideal Airy Disks at design wavelengths for comparison. The discrepancy shows up at the tails of the focusing peaks. The reason is that the incident light source has a finite bandwidth, and the focal spot shifts away from the design focal length when wavelength deviates from the design wavelengths.

5.13 The measured FWHM of the normalized focusing intensity profiles at design wavelengths versus the theoretical FWHM of the normalized focusing intensity profiles of ideal Airy disks at corresponding wavelengths. They agree with each other well at long wavelengths of 610 nm and 650 nm, and show differences at short wavelengths of 490 nm, 510 nm, 540 nm, and 570 nm. The maximum difference is 19% of the theoretical value at the wavelength of 490 nm.

5.14 UV-litography makes the wall of the nanopost slanted as showed in (A) the schematics of the unit cell and in (B) a scanning electron microscope image of a fabricated lens, in (C) the geometry of the FDTD simulations that validated our approximate solver.

5.15 The intensity along the focal axis of a 1 cm diameter lens achromatic for [488, 532, and 558] nm with a numerical aperture of 0.2 shows clear focusing, but also presents spurious foci.
6.1 (A) Top-view figure showing the parameters of the H-shaped meta-atom structure/unit cell. The unit cell has 5 parameters: Lx, Ly, Lx1, Ly1, and the permittivity of the structure. The predictive neural network (PNN) consists of a bi-linear tensor layer, followed by four consecutive fully-connected layers. A ReLU activation function is applied to the output tensor of each layer except for the last one, which is fed to a tanh activation function. The real and imaginary parts of the complex transmission coefficient are predicted using two independent networks that share the same network architecture (B). Details, including four fully-connected hidden layers and five weight arrays, are given in the figure. Courtesy of Sensong An.

6.2 After optimization of a lens of 1 mm diameter with cylindrical symmetry (≈85.5k unit cells with with 2,475 degrees of freedom for 495 independent parameters) for [9.99, 8.57, 7.49, 6.66, 6., 5.45, 5.0] µm, the focal lines for the seven design wavelengths show clear maxima at the desired focal length of 1.485 µm (marked by a yellow line). The presence of multi-foci comes from the multiple constraints. The peak at the focal line at 9.99 µm wavelength clearly translates to a peak at a distance of ≈1750 µm at 8.57 µm wavelength and to a peak at 2000 µm at 7.49 µm wavelength.
7.1 Fully connected neural network (NN) to approximate a low-rank function with three hidden layers containing fifty neurons each (A). (B) Estimates/Error vs targets with 10k training points. On the x-axis, we show the index of the points after sorting the test set w.r.t. to the target value of the real part of the local field. On the y-axis, we show the true real part (test set) in red, estimates from NN in blue dots, and error from NN (orange bar). The test MSE is 0.010. (C) Estimates/Error vs targets with 10k more random training points. The test MSE is 0.008. (D) Estimates/Error vs targets with 10k chosen using active learning. The test MSE is 0.004 which is twice better than the random set of additional points.

7.2 Fractional error of the neural networks’ estimates of the complex transmission on a separate test set. One neural network is trained using our active learning algorithm 2, and the other is trained by adding points randomly to the training set. The two networks start with the same 10000 data points. The active learning approach reduces the variance twice better, which shows early quantitative proof of concept. We also compare to Chebyshev’s interpolation—the state-of-the-art polynomial interpolation method for smooth functions. Active learning reduces the error by 7% by adding 10k points, which is more than twice better than adding 10k random points (2.5% reduction), and Chebyshev’s interpolation with \((7^3 - 6^3) \sim 9k\) additional points (2.8% reduction). 

...
7.3 We performed gradient-based optimization. We obtained the gradient of the surrogate model using backpropagation in the ensemble of neural networks. We show the resulting intensity profile along the focal axis of the lens from the approximate solver based on the surrogate model (orange dots) and its local periodic approximation validation that uses an online solver for the unit cells (solid blue line). The approximate solver based on the surrogate model overestimates the intensity at the focal length, however it computes the profile correctly and has local maxima in the same places as the validation except for one spurious focus annotated by a red arrow. Since the maxima match well, an optimization using the surrogate model also optimizes for solutions of Maxwell’s equation. Note that the discrepancy close to the metasurface (when \( z < 100 \mu m \)) comes from the local periodic approximation of the solver and not from the inaccuracy of the surrogate model.
7.4 We performed gradient-based optimization and obtained the gradient of the mean surrogate model using backpropagation in the ensemble of neural networks for both the mean and the variance of our model. We show the resulting intensity profile along the focal axis of the lens from the approximate solver based on the surrogate model (orange dots) and its local periodic approximation validation that uses an online solver for the unit cells (solid blue line). Again, the approximate solver based on the mean surrogate model overestimates the intensity at the focal length, however it computes the profile correctly and has local maxima in the same places as the validation. Compared to the regular model, the robust model is qualitatively better, because it resolves for all the spurious foci, even the one annotated with a red arrow. Since the maxima match well, an optimization using the surrogate model also optimizes solutions of Maxwell’s equation. Note that the discrepancy close to the metasurface (when $z < 100\mu m$) comes from the local periodic approximation of the solver and not from the inaccuracy of the surrogate model.

B.1 Setup for measurement of the diode laser is collimated on the metalens. The scattered field is then collected by an objective lens, which enables us to measure the focus of a high NA lens. The light is recollimated and then focused on the camera. The measuring equipment is set on a translational piezoelectric stage to enable precise control of its position. The diode laser has a very narrow linewidth that is less than 1nm, while the linewidth of light from the SuperK is larger than 5nm. The bandwidth of light source will affect the depth of focus, as well as point spread function, so the diode laser provides better measurements. The drawback is that we have access to a more limited number of wavelengths.
B.2 The intensity along the focal axis shows clear focusing at the focal spot for the six desired wavelengths which are linearly-spaced in the frequency domain to achieve best achromatic properties in the visible range: [490, 520, 540, 570, 610, and 650] nm. 

(A) The numerical aperture of this lens is 0.1; a diameter of 2 mm corresponds to a focal length of about 1 cm. The lens shows multiple spurious foci along the focal axis. Although the focal spot is clearly demarcated from the spurious foci, the spurious intensities can be bigger than the intensity at the desired focal spot. (B) The numerical aperture of this lens is 0.2; a diameter of 2 mm corresponds to a focal length of about 5 mm. This lens shows similar features as the NA=0.1 case. We plot the focal axis on a larger support compare to Fig. 5.9 to show the multi-foci.

C.1 Coordinate system for the derivation. We consider a waveguide in vacuum and want to optimize the permittivity function $\varepsilon(z)$ to maximize the coupling of incident light to the waveguide mode. Following usual convention for waveguides, the $z$-direction is that of the waveguide.
À ma famille.
Acknowledgments

The African proverb, “it takes a village to raise a child,” resonates with me, and I would say it also took a village to complete my PhD studies.

As my dedication says in French, I dedicate this dissertation to my family: my wife Meredith, my children to come, my parents: Yvelise and Didier, my older sister Isabelle and her family: my brother-in-law Antoine, my niece Jeanne, and my nephew Augustin, my younger sister Mathilde and her partner Arij, my in-laws: Brenda and Mike, and my brother-in-law Chase. I also want to include my extended family, who genuinely rejoice for me when chance smiles at me and sadden for me in cases of adversity. Special thanks to my wife: Ever since I met her at the Museum of Fine Arts in Boston, my life has felt brighter. I am deeply grateful for her love and selflessly involved support. And, special thanks to my parents, who have always supported me in my PhD project, even though the PhD degree is not as valued in France as it is in the United States, where my studies brought me. My mother was instrumental in my application process; she helped me organize my thoughts and proofread my statement of purpose. I consider my father to be my first collaborator. He has followed my research very closely during my whole PhD, bringing a fresh view on it as well as witty ideas. In addition, the support from my sisters via phone calls was instrumental in keeping me grounded. Submitting this
dissertation finishes a project that began 9 years ago; I was working at Airbus in Munich and decided that I wanted to do a PhD at Harvard University. I want to dedicate this thesis especially to the loving memory of my grandparents, Jean-Marie and Jeannine Carret and Fernand and Claudette Pestourie, who are a great part of who I am now. I will always remember their love, their wisdom, and where they came from, and I am grateful and proud of this heritage.

This dissertation would have never happened without my advisors at the Massachusetts Institute of Technology and at Harvard University. I am indebted to Steven G. Johnson, who agreed to take me as his student during a very snowy winter in 2015. I did not know at the time how lucky I was; Steven turned out to be the best primary advisor one can dream of. In the office, he is very professional. He always knew exactly at what step I was in my projects and always had wise advice and answers to the questions I was pondering. Out of the office, he built a kind and thoughtful universe, where the research group becomes a music band during jam sessions, during lunches at hot pot, dim sums, and other exotic restaurants, and during kayaking outings. I am also very grateful to Federico — the man without whom my Harvard studies would probably have not happened. Federico Capasso took me in his laboratory as a visiting student for a few months before I started my PhD. My experience there and the thrill of Federico’s group research environment made me decide to come to Harvard. I also grateful to Federico for his excellent advice and continued support. In addition, I want to thank Chris H. Rycroft who was my mentor since early in my PhD career (and with whom I conducted research and taught a graduate class “Advanced Scientific Computing” as a teaching fellow).

I had the opportunity to find other wonderful mentors in my Harvard journey. I took a class with Susan Murphy for my Master’s in Statistics, which started a wonderful collaboration that I hope will thrive after my PhD. I also thank Stephen Blyth for the opportunity to teach “Quantitative Finance” as a teaching fellow of the Statistics department and L. Mahadevan for his insightful questions during my qualifying exam. I am grateful to individuals at the Institute for Applied Computational Science at Harvard, from which I received secondary field in Computational Science and Engineering: Daniel
Weinstock and Pavlos Protopapas.

I want to quickly mention colleagues with whom I have had the chance to work: at MIT: Fan Wang, David Liu, Adi Pick, Wonseok Shin, Carlos Pérez-Arcencia, Mohammed Benzaouia, and Zin Lin; at Harvard, Romain Blanchard, Tobias Mansuripur, Zhaoyi Li, Elodie Strupiechonski, Patrice Genevet, Guy-Maël de Naurois, Mikhail Kats, Guillaume Aoust, and Benedikt Groever; at UW: Arka Majumdar and Elyas Bayati; at MIT-IBM Watson AI: Youssef Mroueh and Payel Das.

Along my academic career, I have many people to thank. From École Supérieure de Sciences Économiques et Commerciales (ESSEC) I want to thank Marie-Noëlle Koebel for her wise, forward thinking in the matter of a dual degree with École Centrale Paris, Viva Carlon-Pibarot for her support, and Josette Bonelle for her administrative help with the transition to Harvard University. From École Centrale Paris (now CentraleSupélec), I want to warmly thank Pierre Becker for his love for quantum physics, which was contagious, and his fantastic pedagogy; I owe him a significant part for becoming a scientist. I also want to thank Jean-Michel Gillet for his mentorship throughout my student career. During my Master of research in nanoscience, Jean-Jacques Greffet was instrumental in giving me the strong foundation I have been using during my PhD studies with classes at École Polytechnique (X) and the Institute of Optics Graduate School. I completed my Master of research as a visiting scholar at the University of California-Berkeley, where I met Boubakar Kanté, to whom I am grateful for his patient advice and constant support as a mentor and as a friend. I have had the chance to experience multiple institutions during my student career, which I detail here: from Airbus, I want to thank Laurent Bersot and Hervé Buchwalter; from Nanyang Technological University, Alexander Petrovic; from Adionics, I want to give special thanks to Guillaume de Souza for his mentorship; from Laboratoire Kastler-Brossel at École Normale Supérieure, Pierre Cladé; from Forstone Research, Andrey Omelchenko and Luis Ibanez Herrera.

I have also had the chance to have friends who were valuable in supporting in my academic journey: Cédric Neuville, Abdullah Almaatouq, Guillaume Basse, Jules Thierry, Maik Wehmeyer, Senan

I am also grateful to the Fulbright committee from France (especially Arthur Sachs Alumnus Marc-Oliver Timsit) for choosing me as an Arthur Sachs Fellow. I give special thanks to Margot Gill and the fellowship team for making this community thrive on Harvard campus. I want to thank the Harvard Graduate School Leadership Institute (HGSLI) and especially Jen Tutak for her beautiful energy. Among the fellows, I want to thank especially Brian Fuller, Erich Bauch, and Kit Seryak. I am also grateful the Quincy House community, which my wife and I are lucky to belong to: Lee and Deb Gehrke, the wonderful community of tutors, the staff, and students.

To finish, I can think of at least twenty friends (old friends, classmates) who were instrumental in my personal journey. Although this list would belong to an autobiography, I want to acknowledge them here.
From a long view of the history of mankind, seen from, say, ten thousand years from now, there can be little doubt that the most significant event of the 19-th century will be judged as Maxwell’s discovery of the laws of electrodynamics.

Richard Feynman

Introduction

Metasurfaces: definition, motivation and challenges  On the one hand, traditional optics devices rely on propagative properties of the material. As a consequence, these devices are bulky, and often heavy and expensive to assemble together. On the other hand, metasurfaces (also called metalenses for lens applications) rely on the scattering properties of the material and geometry. They are very light and thin and easily manufacturable using the same techniques as the silicon industry (for example, using photolithography). The design of the geometry of a metasurface is key to its per-
formance. In the context of this thesis, metasurfaces are surfaces patterned with subwavelength scale features, aperiodic over length scales much larger than the wavelength. The patterns or unit cells are organized in a mesh of $N$ unit cells in each direction. Each unit cell has a fixed number of parameters, $n$. Metasurfaces are aperiodic, which means that the parameters of each unit cell can be distinct. Therefore, designing a metasurface requires to optimize the $n \times N \times N$ parameters to obtain the best performance. Even the smallest metasurfaces have: $n \in (1, 5)$, and $N \sim 1000$, so it is not uncommon to have to choose millions of parameters. Now my group and I are scaling up $N \sim 10^4$, and $n \sim 10^3$ (billions of parameters). In this thesis we propose to formulate this problem as a large scale optimization problem, whereas previous authors typically use simpler but more limited approaches.

Two very different length scales To make matters worse, simulating Maxwell’s equation for a metasurface represents a real computational challenge. It comes from the fact that metasurfaces present two length scales that differ by orders of magnitude. The smaller length scale is the size of features in the unit cell, on the order of $\sim 10$ nm for applications to visible light. The larger length scale is the diameter of the metasurface. I will present diameters of up to one centimeter in this thesis.

Therefore, the two length scales can be up to 6 orders of magnitude apart. This is very problematic for state-of-the-art/off-the-shelf solvers of Maxwell’s equations because they need to refine their discretization enough to account for the smaller length scale, however, they run out of memory before being able to simulate the whole metasurface. Limiting the design of metasurfaces to what off-the-shelf solvers can solve on a laptop would limit the diameter of the metasurface to a few hundred nanometers. The stark difference in length scales makes simulation intractable with usual solvers because simulating the whole metasurface while resolving the pattern would require a supercomputer with thousands processors for a single known design. Therefore since an optimization requires many solves, it is too prohibitive.
**Research Questions**  My PhD research comprises two themes.

1. How do we simulate metasurfaces? As explained above, simulation of metasurfaces is difficult because of inherent properties of this type of device.

2. How do we find the best design of metasurfaces that will scatter light as desired? Metasurfaces often consist of millions of unit cells, and each unit cell has a discrete number of parameters. How do we determine millions of parameters so that the metasurface scatters light as desired? Determining the parameters one by one would be far too long. Can we update all the parameters together to make steps toward an improving direction?

*How do we simulate metasurfaces?* Prior to my work, the optics community had made a local periodic approximation\(^1\) to determine the phase at the level of the metasurface, thereby overcoming the simulation problem. As the title of this thesis put, the local periodic approximation computes the field of a unit cell by assuming that its neighbors are identical. The assumption transforms the massive simulation into numerous and independent small unit cell simulations. My research has built upon this work to demonstrate the following:

1. The local periodic approximation is accurate up to first order, and higher-order corrections can be computed;\(^2\)

2. A hybrid solver that breaks down the metasurface into multiple small simulations and combines them with an integral equation solver can approximate the scattering of a metasurface very efficiently;\(^3,4\)

3. When the number of degrees of freedom in the unit cell is small, unit cell simulations can be replaced by surrogate models for efficiency. A surrogate model is an approximate model for the field function, which is trained using data (supervised learning). The trade-off is the
following: 1) the surrogate model has to be much faster than a solver of Maxwell’s equations (if it becomes so complex that evaluating it takes a similar time as a solve, it defeats the purpose of a surrogate model), 2) the training time of the model should stay reasonable to achieve accuracy (if the surrogate model needs too much data, such that the training time outweighs the benefit from evaluating the field faster, it defeats the purpose of a surrogate model). I have found two types of surrogate models that work well: Chebyshev interpolation when the unit cell has 1-3 parameters and neural networks where the unit cell has 4 to ~10 parameters (Chap. 6 and 7); 4. Neural network surrogates can be trained effectively using active learning (Chap. 7).

These results brought the simulation time of metasurfaces from supercomputer scale to seconds on a laptop.

How do we find the best design of metasurfaces that will scatter light as desired? Can we update all the parameters together to make steps in an improving direction? Metasurfaces are often constituted of ~100k to millions of meta-atoms, each having up to 10 geometry and material parameters, which adds up to millions of parameters to optimize for a single design. The existing design method for metasurfaces consisted of matching the wavefront at the metasurface with analytic solutions, breaking the large scale optimization problem into numerous independent and small matching problems. For each pattern, the parameters were chosen to match the desired field. This method has two limits: 1) it is required to know the analytical solution of the desired field everywhere above the metasurface, and 2) it cannot do optimization compromises at the device level for multi-objective applications (e.g. optimizing for multiple frequencies), both of which limit the domain of application and the outcome performance.

My PhD research overcame these limits by creating an inverse design framework for metasurfaces,
focused on lens applications. The inverse approach starts with the desired solution. With the least number of assumptions, inverse design sets the design problem as an optimization problem where we maximize our objective—what we care about—to find the optimal set of (millions of) parameters that define our optimal metasurface. Inverse design can naturally extend to more complex optimization problems for multiple angles or frequencies of light. These generally appear as additional constraints in the optimization problem, whereas there were no obvious methods to solve that type of problem using the pre-existing field matching method. Thanks to gradient-based optimization, we show how to successfully change millions of parameters together in a single step towards a better design. Over the past five years, my research has created the following:

1. A minimax optimization framework for balancing multiple metasurface design objectives;³

2. Design and experiment of a lens with increased depth of field in two dimensions;⁵

3. Design and experiment of achromatic lenses in the visible range with large diameters (2 mm and 1 cm) in three dimensions (Chap. 5).

These contributions not only extend the theoretical framework of metasurface design, but they also contribute to actual devices fabricated and measured in laboratory settings, which is the goal of all applied research.

Dissertation outline In Chap. 2, I present the problem of metasurface design more in depth, and more specifically for lens applications. I discuss the state-of-the-art as it was when I started my thesis in 2014, some work done of the local periodic approximation in the uniform case, ² and some extensions of this work. ⁴ In Chap. 3, I present the approximate solver in details and the optimization framework for metasurfaces. This work was published in Optics Express.³ In Chap. 4, I show experimental validation of our optimization framework on a lens application with extended depth of field in two dimensions.⁵ In Chap. 5, I show experimental validations of our optimization framework on
achromatic lens applications in three dimensions with diameter sizes ranging from 2 mm to 1 cm. In Chap. 6, I revisit the surrogate model from Chap. 3, replacing Chebyshev interpolation with a neural network. In Chap. 7, I present an active learning method to train a neural network surrogate model with the least number of simulations needed to reduce the error in the estimate.

List of publications Work in this thesis was submitted to peer-reviewed journals:


• Li, Z., Pestourie, R., Johnson, S.G., and Capasso, F. Inverse design for large-scale and achromatic metalenses. (Manuscript in preparation)

• Pestourie, R., An, S., Zhang, H., and Johnson, S.G. Neural network as surrogate model for inverse design of metasurfaces. (Manuscript in preparation)

• Pestourie, R., Mrouve, Y., Das, P., and Johnson, S.G. Active learning of a surrogate model for Maxwell’s equation. (Manuscript in preparation)

Metalens optimization

Metalens design is the design of metasurfaces for a lens application. In this chapter, we delve into the existing literature about the problem. First, we will focus on the solver with a key approximation—the local periodic approximation (LPA)– which makes computation tractable. We put LPA in the context of the science literature and present a study in the case of uniform surfaces. We then focus on the optimization and show the state of the art in metalens design as of 2016. To finish, we put it all together and show an overview of the optimization framework for metasurfaces of this thesis together.
with some extension work.

2.1 Local Periodic Approximation

2.1.1 LPA in the applied physics literature

The key “locally periodic” assumption is that the unit cell changes sufficiently slowly from one to the next so that we can evaluate its scattered field using a simulation of the unit cell with the periodic boundary condition rather than simulating the whole metasurface, which is (very) costly/untractable.

To our knowledge, the applied physics community was the first to make this approximation for single wavelength lens applications, \(^6,7,8,9\), achromatic lens applications, \(^10,11,12\) and axicons. \(^13\)

2.1.2 LPA in the numerical methods literature

The general idea of breaking a computational domain into pieces and using approximate boundary conditions is known as domain decomposition. Most commonly this method is iterated as in Schwarz iterations (for solving elliptic partial differential equations), each iteration improves the boundary conditions based on the previous solution. \(^14,15\) LPA could be thought as one possible initialization of the approximate boundary conditions of Schwarz iterations as in Algorithm 1. For metalenses, it is sufficient to use LPA boundary condition without additional iteration. But future work may use additional Schwarz iterations to get additional corrections and improve accuracy. Domain decomposition inspired my colleagues to look at neighboring cells. \(^16\)
Result: $X_T$ solution of Maxwell’s equation in each domain $\{\Omega_i\}$
Define a partition of the computational domain $\Omega_i$, $i=1,...,N$;
Define each domain boundary as $\Sigma_i$ (volume around $\Omega_i$);
Compute solution $X_{0,i}$ in $\Omega_i$ with periodic boundary conditions;
for $t=1:T$ do
  Fix the solution in $\Sigma_i$ to be $X_{t-1}$;
  Compute solution $X_{t,i}$ in $\Omega_i$ with $\Sigma_i$ boundary conditions;
end
Algorithm 1: Schwarz iteration algorithm.

2.1.3 Theory work on the local periodic approximation in the uniform surface case

The work briefly summarized here was published in *Optics Express.*

Optical metasurfaces (subwavelength-patterned surfaces described by variable effective surface impedance) are modeled by an approximation akin to ray optics: the reflection or transmission of an incident wave at each point of the surface is computed as if the surface were “locally uniform.” Our research questions were: How quickly does the ray-optics approximation converge as the metasurface’s impedance becomes ever more slowly varying? Can we quickly compute the low-order corrections (both to improve accuracy and to validate ray optics)? How do we compute both far-field and near-field scattering?

In a setting where the surface is along the $x$-direction and $y$-direction is normal to the surface, the ray-optics approximation for a locally uniform surface can be written as follow:

$$u_{0,\text{tot,ff}}(r) = u_{\text{inc}} + \int_{-\infty}^{\infty} \{G_{p}^{\text{ff}}(r|s, o^+)f_{s}^{\text{inc}} - G_{p}^{\text{ff}}(r|s, o^-)f_{s}^{\text{inc}}\} ds, r \in \Omega_+ \cup \Omega_- \quad (2.1)$$
where $\Omega^\pm$ is the domain above(below) the surface, where $u^{\text{inc}}$ is the incident field, and the surface currents $f_+$ and $f_-$ depend explicitly on the incident field and on the impedance and admittance of the metasurface, and the Green’s function $G^{\text{ff}}$ is where the approximation is done, which corresponds with the approximation of the exact Green’s function $G$ in the far-field, and can be thought of as the ray-optics approximation of the total field produced by a point source above (which is case for $O^+$) or below (with is the case for $O^-$) the metasurface. In Ref. 2, we formulated a second-kind integral equation, using an approximate (proto) Green’s function, from where corrections to the ray optics approximations could be computed in the form of a Neumann series. We show that the approximation from Eq. 2.1 corresponds to a zeroth-order locally uniform approximation by presenting a method to compute the solution for any order of accuracy and identifying the approximation with order zero. In Fig. 2.1, we plot the first four order corrections of the field, at an angle of 22.5°. 22.5° is an angle at which the zeroth-order locally uniform approximation starts failing. We also show that the corrections tend toward zero as the metasurface’s impedance becomes ever more slowly varying, which proves a sideways adiabaticity.

The integral equation using the proto Green’s function provides valid approximations for the near and far field. The coupling to the surface mode was achieved using the near field approximation. We noticed that the zeroth order locally uniform approximation could predict surface waves as showed in Fig. 2.2. This fact created a new research direction currently under study about optimizing the coupling of incident field to waveguide modes. This work is further pursued in Appendix A.

2.2 Existing methods of design

There are two approaches to designing optical devices in the literature. The first, mostly used in the applied physics community, uses an analytical approach, and the second is to use optimization or inverse design.
Figure 2.1: High-order corrections to the ray optics approximation.2.1. (a) Geometrical configuration of the problem under consideration which corresponds to an unit amplitude plane-wave impinging at normal incidence on a metasurface that renders a transmitted plane-wave with wave-vector forming an angle of $22.5^\circ$ with respect to the metasurface. (b) and (c): Real part of the total field 0-th and 1st order approximations. (d), (e), (f), (g) and (h): Absolute errors $|u - u_{\text{tot,fit}}^N|$, for $N = 0, 1, 2, 3$ and 4, respectively, in the zeroth (no correction), first, second, third and fourth order corrections to the ray optics approximation $u_{\text{tot,fit}}^N$. The color scales were adjusted according to the maximum error displayed in each one of the figures.

2.2.1 **Analytical method**

For a lens application, the objective field everywhere can be derived analytically. By reciprocity, it is the field that a source would create if placed at the focal point of the lens. It is therefore possible to match the needed field at the level of the metasurface by choosing the parameters of the unit cell that matches the analytic need. In the literature, since the transmission amplitude of the unit cell is roughly constant across parameters when the unit cell is subwavelength, they only match the phase.

For example, for a lens in two dimensions, the target phase of a metasurface along the x-direction
with focal length $f$ is

$$\phi(x) = \left(\frac{2\pi}{\lambda}\right)\left(\sqrt{x^2 + f^2} - f\right).$$

(2.2)

There are several drawbacks to these methods: 1) it is demanding on the unit cell because it requires a phase coverage from 0 to $2\pi$ given the parameter constraints, 2) the matching problem is at the unit cell level while the design problem should be at the metasurface level to capitalize on compromises across unit cells, 3) it is very demanding on the dispersion of the unit cell to use these methods for multiple wavelengths, 4) this approach is very constraining because it constrains the phase everywhere for all design frequencies, so it might constrain the design solution to a subset of design with a lot of poor local optima.

2.2.2 Inverse design

Physically, we can reframe the inverse design problem as finding a generator for a given far field. The solution for this generator is not unique because different near fields could create the same far field. Since the solution is not unique, the inverse problem does not have a unique solution. In practice, we define an optimization problem with an objective function. This approach also has the strength of
finding a (locally) optimal solution, even when the goal is not physically attainable. For example, there are often not enough degrees of freedom in the parameter space to match our ideal need. An inverse design approach ensures that the design we find is at least locally optimal, which is as close as we can get to our ideal given our problem specification. In the case of a lens application, we define the objective function as the intensity of the scattered field at the focal point. A good objective function should have two qualities: 1) it should give the least constraints possible while reaching the desired optimum, and 2) it should capitalize on the mathematical properties of the problem in the most efficient way. We see that the intensity at the focal point is a good objective function because 1) it constrains the field only at one point, and 2) it is differentiable so we can have access to the gradient. Getting the gradient is crucial for performance, because gradient-based off-the-shelf optimization methods are the best methods for non-convex problems involving non-linear constraints. Finding a global optimum is tractable only when the problem has up to ten variables. Since designing a metasurface requires optimizing up to millions of parameters, we can only reach local optima. A local optimum is acceptable as long as the design performs better than the current state of the art.

Before this thesis, some inverse design had been performed in nanophotonics. They introduce a very important concept: using the adjoint problem to get the gradient. The adjoint problem gets the gradient with respect to all parameters in a single simulation of the adjoint problem, therefore obtaining the gradient in two simulations instead of using \( n + 1 \) simulations as a finite difference method would require (for much worse performance). However, prior work was limited to inverse design using brute force solvers: finite difference time domain and finite difference frequency domain. As discussed in Chap. 1, simulation of metasurfaces is currently intractable with a brute force solver, so we cannot apply this strategy directly to the problem of metasurface design.
2.3 “Assume your neighbor is your equal: inverse design in nanophotonic”

Our proposal is to combine LPA with inverse design to design metasurfaces. Assuming LPA holds—“Assume your neighbor is your equal:...”—, we create an approximate design which evaluates the objective function 10⁴ faster than a brute force method for lens application (based on comparison with finite difference frequency domain for a small (30 unit cells) lens application tractable with brute force), and evaluates the gradient with a computational cost similar to a single function evaluation. We use this solver to perform optimization—“Inverse Design in nanophotonics”. The computational framework is formally detailed in Chap. 3 and was published in Ref. 3.

**Online solver versus offline solver**  Assuming LPA holds breaks down the metasurface simulation into multiple smaller independent simulations of the unit cells constituting the metasurface.

If we use a solver to simulate all the unit cells every time we want to simulate the metasurface, then the approximate solver is called *online*. An online solver allows the most freedom in the unit cell. Each pixel of the unit cell can become a parameter, leading to a topology-optimization problem. In Ref. 4, as a follow up of Ref. 3 we explored topology-optimization assuming LPA holds. In Fig. 2.3, we show an example of optimal structure exploiting all the degrees of freedom of the surface. This approach gives generally gives the best results (because it has more degrees of freedom), however it suffers multiple drawbacks: 1) although we can capitalize on the embarrassingly parallel nature of the framework, it still requires a large computer cluster to design a metasurface with this method, 2) even when taking fabrication in consideration, the resulting design is harder to fabricate experimentally.

In contrast, in the case of unit cells defined by few geometry parameters (up to ten), we can train a surrogate model, which will replace the online solver when simulating the metasurface. The surrogate method is called *offline*, because the solver for Maxwell’s equation is used only once to train the surrogate model. This methods is more efficient than the online method, because evaluating the surrogate
Figure 2.3: Monochromatic 3D metalens (NA = 0.37). A few portions of the lens have been magnified for easy viewing; note the scale bars. The lens consists of a single TiO$_2$ layer above the silica substrate. The far field profile is obtained by locally periodic approximation, showing diffraction-limited focusing.

Table 2.1: Benchmark for inverse design in nanophotonics: the first row corresponds to the analytic methods, the second to inverse design with brute force solver, the third to our optimization framework. The first column corresponds to gradient based optimization, the second to the ability to simulate a metasurface, the third to the ability to design for objective that do not have an analytic solution, the fourth to the capacity to handle multiple constraints optimization. Our approach (Inv. Design/ LPA) fills the gap for inverse design of metasurfaces compared to the state of the art.

<table>
<thead>
<tr>
<th>Gradient</th>
<th>Metasurface</th>
<th>Non-analytic</th>
<th>Multi-Const.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Inv. Design/ BF</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Inv. Design/ LPA</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

model takes much less computational effort than solving Maxwell’s equations. We show two types of surrogate models in this thesis: 1) a polynomial presented in Chap. 3, 2) a neural network presented in
both Chap. 6 and 7. In Chap. 4 and 5, we present applications from actual experiments, which were optimized in at most a few hours using our offline optimization framework on a laptop.

**Overlapping domain and online vs offline solver** The overlapping domain method consists of computing the field in a unit cell, while including the neighboring unit cells in the local simulation. We can then replace the periodic boundary condition by an absorbing boundary layer since the near-field effect of the neighboring cells is already taken into account by physically adding them, and neglecting long range scattering which is appropriate for a lens. The overlapping domain technique was inspired by analogous methods in domain decomposition (Schwarz iteration). This method has proven to be useful in cases where LPA breaks down, however it dramatically increases the number of parameters for the Maxwell’s equation solver. In two dimensions, the number of parameters is multiplied by 3 (the unit cell of interest and its two neighbors). In three dimensions, the number of parameters is multiplied by 9 (because there are now eight neighbors in the mesh). This tends to favor an online solver approach where the number of parameters does not really matter. However, surrogate methods presented in Chap. 6 and 7 can support overlapping domain techniques when there are only one or two parameters per unit cells.

In Table 2.1, we show a benchmark of methods for inverse design in nanophotonics. Our approach (Inverse Design/ LPA) fills the gap for inverse design of metasurfaces compared to the state of the art.
Inverse design of large-area metasurfaces

3.1 Introduction and motivation

In this chapter, we present and validate a fast method for optimization-based “inverse design” of large (hundreds of wavelengths λ) aperiodic metasurfaces for wavefront shaping.\textsuperscript{24,25,26,27,28} incorporating both scattered amplitude and phase for multiple incident λ and angles. Previous methods either op-

\textsuperscript{1} Work published in Pestourie et al., 2018 \textsuperscript{3}
timized the full Maxwell equations, \textsuperscript{29,30,31,32,33} (which is infeasible for large surfaces), were restricted to weakly coupled scatterers, \textsuperscript{34} or started with a desired scattered phase and tried to design a corresponding metasurface unit cell, as in Ref. \textsuperscript{6,7,8,9,10,11,12,13,35,36}—henceforth referred to as “analytical design articles,” (but if attainable unit cells fail to exactly match the desired \(\lambda\)-dependent phase there was no systematic way to choose the best compromise). In contrast, our approach starts with a family of manufacturable unit cells and directly optimizes an aperiodic composition for the desired field pattern by a fast approximate model, automatically finding the best compromise for the given constraints. Whereas phase-design methods typically assume that the desired scattered field is known everywhere (analytical design articles), our approach allows one to specify the field objective only in regions of interest.

As outlined in Fig. 3.1, given exact scattering calculations for small metasurface unit cells (Sec. 3.2.1 and Fig. 3.2), we build an approximate convolutional model of an arbitrary metasurface (Sec. 3.2.2) that can then be optimized (Sec. 3.3) rapidly (seconds to find the optimum for a 200-\(\lambda\) aperiodic surface with hundreds of parameters) using two different objective functions. We validate the optimized design with a brute-force Maxwell solver (Sec. 3.3.1) and we find excellent quantitative agreement (Fig. 3.4) even for rapidly varying aperiodic surfaces that challenge the assumptions of our model. We present example designs (Sec. 3.4) for a multi-wavelength optimization (Fig. 3.5), a wavelength demultiplexer (Fig. 3.6), and an multi-angle optimization (Fig. 3.7). Our approach is not limited to true “metasurfaces” whose features are small enough to mimic effective-impedance surfaces: we show that it even works well for large-period microstructures that scatter multiple diffracted beams. Indeed, our method is easily extensible to incorporate multiple diffraction coefficients, multi-layer/multi-parameter unit cells, multiple polarizations, and other complications (Sec. 3.5 and Fig. 3.8).
3.2 **Locally periodic approximation**

The key to “metasurface” design is to be able to quickly calculate the transmitted/reflected field for a large-area structure, possibly thousands of wavelengths in diameter—too large to solve the full Maxwell equations without some simplifying assumption. Similar to analytical design articles, the central approximation of our approach is to assume that the metasurface is *locally periodic*: the scattering in any small region is almost the same as the scattering from a periodic surface. The use of periodic calcu-
lations to compute the specular reflection phase only, typically discarding amplitudes and additional diffracted orders, has sometimes been called a “local phase approximation” \(^{46,47}\) (Contrast this with the regime of scalar diffraction theory, \(^{48,49}\) valid for period \(\Lambda \gg\) wavelength \(\lambda\), in which the surface is treated as locally *uniform*, separately computing the transmission coefficient at each point on the surface.) In a separate paper, \(^2\) we develop the rigorous foundations and convergence rates of a related approximation, along with higher-order corrections, but here (similar to analytical design articles) we will simply perform brute-force Maxwell simulations at the end (Sec. 3.3) to validate our designs. (In fact, we will find in Fig. 3.4 that the locally periodic approximation gives excellent agreement with full simulations even for surfaces where the unit cell is rapidly varying in some regions.) Unlike previous authors who calculated only the scattered phase and not the total scattered field (analytical design articles), we employ both amplitude and phase information to formulate a complete approximate solver (scattered field for any given incident field) that can be used to optimize the metasurface for arbitrary “objective” functions of the field. Not only does this approximate solver enable very general optimization, it also allows us to evaluate the optimized metasurface for different (non-optimized) incident fields (e.g. the wavelength sensitivity computed in Sec. 3.4 for the multi-wavelength lens). As discussed in Sec. 3.5, our approach is also easily extensible to a ”non-metasurface” regime in which there are multiple diffracted beams from a large-period surface, as well as to computing near fields (via the scattering coefficients of the evanescent waves). In this section, we explain in detail how the locally periodic approximation allows us to compute the total scattered/expected field (at any point in space) for any incident wave.

To make it easier to understand our approach, it is helpful to consider a specific example of a two-dimensional “metasurface” unit cell, based on \(^{12}\) TiO\(_2\) pillars on top of a silicon dioxide substrate, as shown in Fig. 3.2. The height of the pillar is fixed to 600 nm, the period \(a\) is fixed to 235 nm, and the pillar width varies: \(p \in [50, a - 50] \) nm (imposing a minimum feature size of 50 nm for practical fabrication). One could easily add more parameters and/or constraints, as discussed in Sec. 3.6. Given
this unit cell, an aperiodic metasurface is formed by taking a group of such unit cells with independent parameters and juxtaposing them next to each other.

Our goal is to compute the scattered (transmitted or reflected) field for such an aperiodic surface, for any given incident wave (e.g. a planewave or Gaussian beam), given only the exact Maxwell solutions for scattering of planewaves by periodic surfaces of the different pillar widths. In this paper, we consider only incident propagating (not evanescent) waves, but in another paper \textsuperscript{2} we show that a similar approach can be extended to evanescent fields as well. The key “locally periodic” assumption is that the pillar width (the unit cell) changes sufficiently slowly from one pillar to the next. (This assumption is rigorously quantified in \textsuperscript{2}, is validated numerically in Sec. 3.4, and it turns out that we even obtain good accuracy when there are sudden changes in pillar width at a few locations.) As mentioned above, this assumption is similar in spirit to other metasurface work (analytical design articles), where it was found to work well for a wide variety of metasurface designs; the main contribution of this paper is to couple the locally periodic approximation to general optimization tools and near-to-far-field transformations. Of course, this approximation can break down for devices requiring extremely rapid surface variations such as diffraction to nearly glancing angles,\textsuperscript{50-2} although it can be generalized by including a next-order correction,\textsuperscript{2} but this is not a problem for the moderate-NA lens-like applications considered in this paper.

3.2.1 Periodic sub-problems

In Fig. 3.2(left) is shown the fundamental assumption of our approach. For each unit cell of the aperiodic structure, we approximate the field in a plane/line just above the unit cell by the solution for the equivalent periodic structure. Three examples are highlighted corresponding to three different parameters of the unit cell. When the period of the unit cell is subwavelength, the zeroth diffractive order is the only propagating wave.\textsuperscript{51} Therefore, if we are interested only in the far field, we can make an additional approximation: we replace the scattered field by its zeroth Fourier component which
Figure 3.2: Left: an arbitrary aperiodic metasurface (top) is approximated by solving a set of periodic scattering problems (bottom), one for each unit cell, to obtain the scattered field just above the surface (horizontal line segments). Right: 0th diffracted-order amplitude (top) and phase (bottom) of periodic subproblems as a function of the pillar width. This is precomputed for several widths (markers) and interpolated as needed.

is simply the average of the field on the plane just above the pillar. Given this approximate field just above the surface, in Sec. 3.2.2 we construct an approximate field everywhere above the surface. In Sec. 3.5, we go beyond the zeroth-order (specular) approximation by including additional diffractive
orders.

We will consider periodic structures with hundreds of different pillar widths (with a fixed period), but we would like to avoid having to do hundreds of unit-cell calculations. We can take advantage of the fact that the scattered fields are smooth functions of the pillar width by solving the scattering function for a few widths and then interpolating to any other widths.

Given a smooth function \( f(p) \) of some parameter \( p_1 \leq p \leq p_2 \), Chebyshev’s methods evaluate \( f(p) \) at a few special points \( p \) and construct a polynomial approximation \( \tilde{f}(p) \) that can be used to rapidly evaluate the \( f(p) \) with exponentially good accuracy. This can be extended to multiple parameters using products of Chebyshev’s polynomials or by more sophisticated methods such as sparse grids for more than 3–4 parameters. In this way, we only need to solve the unit-cell Maxwell problem a few times to obtain our polynomial approximant \( \tilde{f}(p) \), which is then evaluated, along with its derivative, many times during optimization. In particular Fig. 3.2(right) shows the amplitude, the and the phase of the zeroth Fourier coefficient of the transmitted field versus the pillar width. We evaluate this coefficient for 21 different widths (at Chebyshev points), and can then interpolate to high accuracy using Chebyshev polynomials. Here, we use the finite-difference frequency-domain (FDFD) method with perfectly matched layer (PML) absorbing boundaries, but any other computational method for periodic Maxwell problems would work as well.

3.2.2 Green’s functions and the equivalence principle

Once the fields are known in a plane above the metasurface, we can obtain the fields everywhere above the metasurface using the principle of equivalence, also known as a near-to-far-field transformation: the fields in the \( y = y_0 \) plane can be treated as equivalent current sources that generate the
fields everywhere else. These equivalent electric (J) and magnetic (K) current densities are defined by:

$$\begin{bmatrix} J \\ K \end{bmatrix} = \delta(y - y_0) \begin{bmatrix} \hat{n} \times H \\ -\hat{n} \times E \end{bmatrix}$$  \hspace{1cm} (3.1)$$

where $\hat{n} = \hat{y}$ is the surface unit-normal vector and the delta function implies that these are surface currents on the plane $y = y_0$.

A further simplification is possible if we only care to compute the fields above the surface. The currents (3.1) produce the desired scattered fields above the plane and zero fields below the plane, and this means that the same fields above are produced if we add or subtract the mirror-image currents (which produce fields below and zero above). Subtracting the mirror-image sources, however, cancels the J term and leaves only the K current (a pseudovector under mirror flips). This allows us to use only $\hat{n} \times E$ sources arising from the electric field computed by the locally periodic approximation in section 3.2.1. As explained in section 3.2.1, we can further approximate the E field by its average in each unit-cell calculation for subwavelength periods, since this gives the far-field diffracted order.

Given these equivalent currents, or their approximation by far-field locally periodic calculations, the electric (or magnetic) fields at any point $x$ above the surface can be computed by integrating along with the Maxwell Green’s function (the field at $x$ from a source at $x'$). For our two-dimensional model problem ($xy$ plane) with the $E_z$ polarization, where we only have a current $K_x(x') = -E_z(x') \delta(y - y_0)$ and we let $G(x, x')$ denote the relevant component of the Green’s tensor [$E_z(x)$ from $K_x(x')$], this integral takes the form

$$E_z(x) = -\int_{\text{surface}} G(x, x') E_z(x') \, dx'$$  \hspace{1cm} (3.2)$$

where $G$ is a Hankel function $G(x, x') = -\frac{ik}{4} H_1^{(1)}(kr)\hat{n} \cdot \frac{r}{r}$, where $k = 2\pi/\lambda$, $r = x - x'$, and $r = |r|$. For a finite metasurface with an infinite silica substrate, we use a standard “windowing” method to truncate this integral accurately to a finite region.
This equivalent-currents formulation is exact if the true aperiodic $E_z$ field is used for the $K_x$ source term, and in section 3.3 we find that it has excellent accuracy with the locally periodic approximation for typical metasurface designs. (A related approximation is made in scalar diffraction theory, where the locally uniform approximate scattered fields can be thought of as sources producing fields everywhere else, further approximated in the far field by e.g. the Fraunhofer diffraction theory.)

### 3.3 Metasurface inverse design methods

The previous section gives us a fast way to solve the forward problem for the scattered field above a given metasurface. In this section, we see how we use it to solve the inverse problem, i.e. find the parameters of a metasurface to produce a desired scattered field. We solve it as an optimization problem: we minimize or maximize an objective function of the unit cell parameters subject to some constraints. Given an efficient way to compute any objective function and its gradient, there are a wide-variety of well-known optimization methods that can be applied; we use the “CCSA-MMA” algorithm via a free-software implementation. To avoid getting trapped in poor local optima, we use a technique called successive refinement: We successively double the number of degrees of freedom, using the optimized coarser structures as starting points for optimization of the finer structures. (The result was not very sensitive to the starting parameter guess; we simply started each parameter in the middle of its allowed range.) What objective function should we optimize? In Sec. 3.3.1 we consider an objective function similar to previous work, which matches the field just above the metasurface with the desired field. In Sec. 3.3.2 we optimize more general functions of the scattered field, e.g. the intensity at a single focal point, which is more flexible when only partial information is known about the desired field. In both cases, in this section we will design a simple lens structure that we will validate using brute force simulations. In Sec. 3.4, we will consider more difficult design problems. In Sec. 3.3.3 we generalize our approach to multiple frequencies and angles of incidence via a maxmin formulation.
Figure 3.3: Bottom: geometry of a metasurface designed for a 5-degree incident plane wave of wavelength 532 nm and focal length 14.7 μm (numerical aperture of 0.3) using the wavefront method. This design produces a field with the needed phase (middle). Top: $|E|^2$ intensity plot shows focusing to the target focal spot.

3.3.1 Optimizing the wavefront

When the exact desired field is known everywhere above the metasurface, as in lens design (analytical design articles) and other wavefront shaping problems, by the equivalence principle it is sufficient
to produce this field on a plane just above the surface. Since the approximate scattered fields \( s(p) \) just above the surface (the \( E_z \) produced for a given metasurface parameter \( p \)) are given by the locally periodic approximation in Sec. 3.2.1, we can directly minimize the difference between this \( s \) and the desired field \( a(x) \exp \{ i\varphi(x) \} \):

\[
\text{minimize } \int |s(p(x)) - s_0 a(x) \exp \{ i\varphi(x) \} + i\varphi_0|^2 dx, \tag{3.3}
\]

where \( s_0 \) and \( \varphi_0 \) are an unknown overall amplitude/phase and \( p(x) \) describes the metasurface parameters along the surface. This approach eliminates the need for any Green's function integral (Sec. 3.2.2) to obtain the field elsewhere.

For a lens application, typically \( a(x) = 1 \) and all of the information is in the desired phase \( \varphi(x) \).\(^{10}\)

A closely related approach was used for metalens design in several previous works (analytical design...
There, since both $a(x)$ and the locally periodic far-field $|E_z|$ were approximately constant, the amplitude was ignored and they simply attempted to match the desired phase. If this phase can be matched exactly in a given unit cell by tuning its parameter $p(x)$ (e.g. pillar width), then no explicit optimization formulation is needed, but an optimization-based approach is more flexible at balancing tradeoffs in cases where the desired $a \exp i\phi$ cannot be exactly obtained, especially in multi-frequency problems (Sec. 3.4.1). A phase-based optimization approach was directly employed in 36 for topology optimization of a small area (no locally periodic approximation).

For example, in Fig. 3.3 we minimize equation (3.3) for a single-frequency $\lambda = 532$ nm lens design problem: we focus an incident planewave at a 5-degree angle on a focal point 14.7 $\mu$m from the surface, using the target phase $\phi(x)$ from 10. We optimize over piecewise-constant parameters, effectively one parameter per unit cell, with a standard optimization algorithm utilizing analytically computed gradients of the objective function with respect to the parameters. Starting the optimization from a constant-$p$ initial guess was sufficient to obtain a local minimum with excellent performance shown in Fig. 3.3. (This 40 unit-cell optimization required $< 100$ ms on a laptop.) At the top is the $|E_z|^2$ intensity plot computed with our approximate solver (Sec. 3.2.2). Below this is shown the 96% match between the desired and obtained fields (from the locally periodic approximation) just above the metasurface. At the bottom is shown the optimized metasurface geometry, which is mostly slowly varying but has sudden jumps in the pillar widths when the desired phase passes through $2\pi$.

In Fig. 3.4, the locally periodic approximate solver (left) is compared to a brute-force surface-integral equation (SIE) Maxwell solver for this optimized solution, showing good quantitative agreement. More precisely, at right we compare the computed intensities $|E_z(x, y)|^2$ for several separations $y$ from the surface. On the focal line, the mean squared difference between the solutions divided by the mean squared intensity is only 0.3%, validating our locally periodic approximation. The errors increase as one approaches the surface because of effects that decay with distance—scattered waves (intensity $\sim 1/y$) from sudden jumps in the pillar width (which violate the locally periodic approx-
imation) along with evanescent fields that we neglected in our far-field approximation—combined with the fact that small errors are more apparent in low-intensity regions far from the focal point.

3.3.2 Optimizing arbitrary functions of the field

Alternatively, since Sec. 3.2.2 allows us to compute the approximate field anywhere above a metasurface, we can optimize any function of this field. This is especially useful if the desired field is only partially known: perhaps one cares about the field in some regions but not others, or is interested in amplitude but not phase. In particular, here we approach the lens-design problem by directly maximizing the intensity $|E_z(x)|^2$ at a single focal point $x$, which can be rapidly computed by a single integral (3.2) of the locally periodic surface fields. As in the previous sections, we used standard optimization techniques\textsuperscript{18} with an analytically computed gradient (essentially via an adjoint method\textsuperscript{72}), and the optimized structure for 40 unit cells was found in $< 1$ s on a laptop (whereas our brute-force solver was about $10^5$ times slower). A comparison of the two methods when the period is not sub-wavelength appears in Sec. 3.5.

3.3.3 Max–min optimization for a combination of multiple objectives

Many design problems involve a combination of multiple objectives: maximizing performance at different wavelengths, angles, and/or focal spots, for example. One common way to do this is a max–min formulation: we optimize the worst objective

$$\max_{\text{parameters}} \left[ \min_{l \in \text{wavelengths}} \text{objective}(\text{parameters}, l) \right].$$

(For example, in Sec. 3.4.1, the “objective” function for an RGB lens is the intensity at the focal spot, and max–min optimization means that we try to maximize the lowest intensity across the three design wavelengths.) Although the expression $[\cdots]$ being maximized is no longer differentiable, which
would make the most efficient high-dimensional optimization methods inapplicable, it can be transformed into an equivalent differentiable problem\textsuperscript{73}

\[
\max_{t, \text{params}} \ t
\]

subject to \( t \leq \text{objective}(\text{params}, \lambda) \) for \( \lambda \in \text{wavelengths} \).

where \( t \in \mathbb{R} \) is a new “dummy” optimization parameter, and params is an abbreviation for parameters. Assuming that the original objective function is differentiable, we can now use a standard nonlinear constrained-optimization algorithm.\textsuperscript{18} In particular, the CCSA-MMA algorithm\textsuperscript{18} only requires us to supply the functions \( \text{objective}(\text{params}, \lambda) \), \( t \) and their gradients (with respect to \( t \) and the parameters) in order to solve the local-optimization problem. Efficient gradient formulas for our cost functions from Sec. 3.3.1 and Sec. 3.3.2 are given in Appendix A.

We will show examples of such optimization problems in Sec. 3.4, where we will use max–min to optimize for multiple frequencies (Fig. 3.5 and Fig. 3.6) or angles (Fig. 3.7).

### 3.4 Applications: RGB lens, demultiplexer, and angle-insensitive lens

In this section, we show some larger and more interesting design problems that can be solved by our methods from the previous section. We still use the same TiO\(_2\) pillar unit cells as in Sec. 3.2, but now we consider metasurfaces consisting of 1000 unit cells, combining multiple frequencies and/or angles, and we could solve the resulting optimization problems in a few minutes on a laptop. In particular, we consider three applications: a lens which has the \textit{same} focal spot for RGB (red, green, blue) wavelengths, a demultiplexer that focuses RGB wavelengths at three \textit{different} focal spots, and a lens that focuses four incident \textit{angles} at the same wavelength to the same focal spot. We will also show that our methods are suitable for sensitivity analyses with respect to wavelengths or angles, by evaluating designs at non-optimized inputs using our fast (locally periodic) solver.
3.4.1 Max–min RGB (red, green, blue) focusing

Here, we use the max–min method of Sec. 3.3.3 to focus normally incident plane waves of three different wavelengths—480 nm (blue), 530 nm (green), and 650 nm (red)—on a single focal spot, by maximizing the minimum (worst) intensity at that spot for all three wavelengths. The diameter of the lens is 235 microns (1000 unit cells), and the focal length is 350.6 microns, which corresponds to a numerical aperture of 0.3.

At the bottom of Fig. 3.5 is shown the intensity on the focal line for all three wavelengths, demonstrating nearly diffraction-limited focusing (RGB half-maximum widths of 975, 997, and 850 nm, respectively). In Fig. 3.5(middle), we evaluate our optimized design along the focal axis (a fixed $x = 0$) versus distance $y$ from the surface and versus wavelength across the visible spectrum, in order to show the wavelength sensitivity of our RGB design. This plot reveals that the optimized design is actually producing three different focal spots (local intensity maxima) on the focal axis for every wavelength, and at each of the RGB wavelengths a different spot is brought to the 350.6 $\mu$m target. At this target focal spot, the intensity $|E_z|^2$ is plotted versus wavelength in Fig. 3.5(top), showing the narrowband nature of the RGB focus. The ability of our approximate solver to rapidly evaluate the performance of the design with many different (non-optimized) inputs ($<100$ ms each) is a powerful tool for characterizing and understanding the metasurface.

3.4.2 Demultiplexer

Here, we design a demultiplexer that focuses normally incident plane waves of three different wavelengths (RGB again) at three different points, which are sixty microns laterally ($x$) apart from each other on the same focal plane (again 350.6 $\mu$m from the surface, a numerical aperture of 0.3). As above, we use the max–min formulation from Sec. 3.3.3 to maximize the worst case intensity at the focal spots.
Figure 3.5: Bottom: the focal line of the scattered field for the three target wavelengths (blue, green, and red) show a clear focusing on the target focal axis. Middle: sensitivity plot for the focal length with respect to the wavelength show chromatic aberration, and each wavelength objective creates a "spurious focus" (local maximum along along the focal axis) on the focal axis at other wavelengths. The red spots represent the foci for each wavelength, and we clearly see chromatic aberration. Top: intensity at the target spot vs. wavelength.

In Fig. 3.6(top), we show the field intensities in the vicinity of the three focal spots for the RGB wavelengths, and in Fig. 3.6(bottom) we plot the corresponding intensities along the focal line $y = 350.6 \mu m$. The focal spots for the two side focal points are tilted outward from the focal axis, which makes sense because they required off-axis focusing relative to the center of the metasurface. As in Sec. 3.4.1, we attain nearly diffraction-limited RGB foci half widths of 825, 785, and 795 nm, respectively.
Figure 3.6: Bottom: focal lines for the three target wavelengths (blue, green and red) focus on points sixty microns apart. Top: the field produced by our design focuses on the desired foci, the high-intensity regions for blue (left) and red (right) are tilted because their foci are off-axis.

3.4.3 Max–min multi-angle focus

Our last application is a metasurface focusing incident plane waves coming at four different angles of incidence (normal $0^\circ$, $3^\circ$, $6^\circ$, and $9^\circ$) at the same focal point for the wavelength $532$ nm, inspired by earlier topology-optimization work. As in the previous sections, we target a focal length of $350.6 \mu m$ (numerical aperture 0.3), and use the max–min formulation of Sec. 3.3.3 to maximize the worst-case focal-point intensity.

Fig. 3.7(right) shows the field intensities in the vicinity of the target focal spot for the four angles, exhibiting an unsurprising “tilt” proportional to the angle of incidence. As in the previous sections, the spots are nearly diffraction limited (half widths of $787, 787, 807,$ and $724$ nm). Fig. 3.7(left) shows the corresponding intensities on the focal plane $y = 350.6 \mu m$ versus $x$. This plot shows that, in addition to a peak at the target point $x = 0$, the metasurface produces three auxiliary side peaks. (Prelimi-
nary work indicates that, similar to [36], these auxiliary peaks can be mostly eliminated by redesigning the unit cell via additional parameters; we will address this in a future manuscript.) That is, much like in Fig. 3.5, the metasurface is creating four focal spots, such that at each angle of incidence a different focal spot is brought to the $x = 0$ target point. The complex surface design and resulting transmitted field here would be very difficult to reproduce without large-scale optimization.
3.5 Beyond subwavelength periods

The term “metasurface” should strictly apply only to deeply subwavelength structures that can be accurately described by an effective surface impedance/admittance or similar, and most previous work operated in a subwavelength regime (analytical design articles). Conversely, when the period is larger than the wavelength, additional diffracted waves appear in the far field that cannot be described by a uniform effective medium or by a single Fourier coefficient. Nevertheless, if the unit cells are mostly slowly varying it should still be valid to describe the surface by a locally periodic approximation (analogous to the adiabatic theorem for propagation through nearly periodic media) to approximate the field just above the surface and hence the field everywhere as in Sec. 3.2. When we solve the local periodic problems in non-subwavelength structures we can no longer retain only the 0th order Fourier coefficient, but instead we must retain either the full $E_z$ field on the surface or, for far-field calculations, the Fourier coefficients corresponding to all of the non-evanescent diffracted orders.

In Fig. 3.8 we show a single-wavelength ($\lambda = 532$ nm) lens design for a period of $800$ nm $> \lambda$, so that even a periodic surface produces two additional diffracted orders $\pm 1$ in addition to the 0th-order “specular” transmission. (Other than the period, the structure is the same TiO$_2$ pillar geometry considered in the previous sections, we use normal incidence, and design for a focal length of 48.6 $\mu$m with 40 unit cells similar to Sec. 3.2.) We considered both the wavefront and the intensity optimization approaches, and validated against a brute-force Maxwell solution as in Sec. 3.3. Since the additional diffracted orders propagate at oblique angles, they have little influence on the focal intensity if the lens is designed to focus the 0th-order (specular) transmitted wave sufficiently far from the surface, so we carry out the inverse design using only the 0th-order term in the approximate model.

The results in Fig. 3.8 show that the intensity method still produces an excellent (near diffraction-limited) focal spot with high intensity that agrees well with the brute-force validation, whereas the
wavefront optimization produces a much weaker focus that agrees poorly with the validation. In both cases, the brute-force calculation and the approximate solver (which includes also the diffractive orders \(\pm 1\)) clearly show the additional diffracted orders scattering to oblique angles that have low amplitude at the focal spot. One major problem with the wavefront approach in this geometry is that varying the pillar width in this case changes the amplitude from 1 to 0.2, very different from the constant amplitude \(\approx 1\) in the subwavelength case. The best phase match corresponds to a weak efficiency, whereas the intensity method can compensate by utilizing both amplitude and phase variations. The resulting lens designs shown in Fig. 3.8(left) correspondingly have an average amplitude twice bigger for the intensity approach than for the wavefront approach (0.8 vs 0.4). Another challenge of non-subwavelength structures, which would become more acute for larger-aperture lenses, is that large-period gratings with only a small number of parameters per unit cell cannot easily implement the rapid variations in phase that are called for by large lenses. There are too few parameters to fit the complex intra-cell phase variation.

### 3.6 Concluding remarks

We believe that our locally periodic inverse-design approach represents a powerful extension to the ideas in previous work, allowing one to balance competing tradeoffs in wavefront design, optimize arbitrary functions of the scattered field (e.g. intensity in selected regions), evaluate parameter sensitivity, design for robustness to uncertainties, and to go beyond the regime of subwavelength structures and far-field designs. A similar max–min formulation can be used to implement a standard robust optimization method to account for manufacturing uncertainty.\(^{69,78}\) Our approximate solver remains orders of magnitude faster than optimization methods based on full Maxwell solvers, allowing it to scale to aperiodic structures hundreds or thousands of wavelengths in diameter while retaining acceptable accuracy for typical designs. We find that complex behaviors can be designed even from very simple
Figure 3.8: Bottom: the geometry (left) from intensity optimization shows big variations in the width of the pillar, and produce good focusing when simulated with a brute force simulation (right), or our locally periodic solver (middle) which includes the diffractive orders $\pm 1$. Top: the geometry (left) from wavefront optimization shows poor focusing both using our locally periodic solver (middle) or a brute force calculation (right). All the intensity plots have the same color scale.

unit cells without plasmonic resonances, and without operating in deeply subwavelength regimes.

This paper presented a proof of concept and validation of the approach, and opens up many fu-
ture possibilities. We are currently working on extension to the design of 3d surfaces and vector fields, and believe such problems to be tractable with a few hours of computation (rather than the few minutes required here for 2d inverse problems). We can easily extend our inverse design from a single parameter per unit cell to multiple parameters per cell. With a few \( \lesssim 10 \) parameters, one can use a similar library-based approach via multidimensional interpolation, for which the main limitation is the number \( N \) of unit-cell calculations that need to be solved beforehand in order to build the interpolation library. The simplest method is a tensor product of Chebyshev polynomials, which is practical for at most 2–3 parameters because \( N \) grows exponentially with the number of parameters. Polynomial scaling of \( N \) can be achieved by sparse-grid methods or neural networks. To handle hundreds or thousands of parameters per unit cell for topology optimization, the library approach must be abandoned in favor of directly solving Maxwell’s equations in every metasurface unit cell for each optimization iteration (still via the locally periodic approximation). In this case, the cost is essentially independent of the number of parameters and scales linearly with the number of unit cells, which can be solved in parallel; we have successfully optimized metasurfaces with > 1000 parameters per unit cell in this way and are currently preparing a manuscript on those results. Multiple parameters per unit cell could describe more complicated surface patterns (e.g. the V-shaped antennas of 27), but also includes the possibility of multi-layer patterns (e.g. stacked gratings). Additional degrees of freedom could prove crucial for obtaining truly wide-bandwidth devices, coupling multiple polarizations, minimizing unwanted reflections, and so on.

The ability to design non-subwavelength surface patterns (but still far from the \( \gg \lambda \) regime of scalar diffraction theory) could prove useful for a variety of applications, starting with designs for short wavelengths (e.g. near UV) where subwavelength fabrication is difficult. The additional diffracted orders of large-period structures may also become useful for near-field focusing and related design problems or for focusing a single incident beam at multiple spots.

Another interesting direction to explore would be further development of the theory of nearly pe-
periodic structures and locally periodic approximations. In a companion work,² we develop a rigorous theory of slowly varying (nearly uniform) structures, and show that the analogous “locally uniform” approximation appears as the 0th-order term in a convergent series of integral corrections. A corresponding rigorous theory of higher-order corrections to the locally periodic approximation, analogous to coupled-mode expansions for propagation through nearly periodic media,⁷⁴ along with efficient numerical methods to obtain corrections, is an important goal for the theory of metasurfaces. A closely related problem is coupling radiation to and from guided modes by nearly-periodic surfaces, a version of which is solved in ². In another paper,² we have recently shown that similar local approximations can indeed be used to compute both near fields and coupling to guided waves.
It doesn’t matter how beautiful your theory is, it doesn’t matter how smart you are. If it doesn’t agree with experiment, it’s wrong.

Richard Feynman

Lens with extended depth of field

This work comes from a manuscript in preparation: Elyas Bayati*, Raphaël Pestourie*, Shane Colburn, Zin Lin, Steven G. Johnson, Arka Majumdar ”Inverse designed metalenses with extended depth of focus”.5
4.1 Introduction

Sub-wavelength diffractive optics, also known as metasurfaces, have generated strong interest in recent years due to their ultrathin nature and ability to arbitrarily manipulate optical wavefronts. By virtue of a large number of scatterers, each capable of shaping the phase, amplitude or polarization of light, metasurfaces can create ultra-compact optical elements, including lenses, polarization optics, axicons, and free-form optics. While metasurfaces provide an extremely large number of degrees of freedom to design complex optical functions, our intuition often fails to harness all these degrees of freedom. One promising solution is to employ computational techniques to design metasurfaces, where the design process starts from the desired functionality, and the scatterers are designed based on a specified figure of merit. Such design methodologies, often referred to as inverse design, have been employed to design high efficiency periodic gratings, monochromatic lenses, and achromatic lenses. Till date, however, no experimental demonstration of inverse designed aperiodic metasurfaces has been reported that exhibit superior performance to an optical element designed via intuition, including traditional refractive optics. For example, recently demonstrated inverse-designed cylindrical lenses exhibit high efficiency, though the efficiency is not higher than that of a traditional refractive lens.

Extended depth of focus (EDOF) lenses represent an important class of optical elements, specifically for their utility in microscopy and computational imaging. This class of lenses differs from an ordinary lens, as the point spread function (PSF) of the lens remains the same over a large distance along the optical axis. EDOF lenses not only enable novel functionalities, such as bringing objects at different distances away from the lens into focus but also alleviate the stringent requirements on aligning lenses on top of a sensor. While for an ordinary lens the gap between the lens and the sensor plane needs to be very close to the focal length, with EDOF lenses, the gap can vary to a degree without sacrificing the performance of the imaging system. In recent years, metasurfaces have been employed
to implement EDOF lenses. Existing EDOF lenses, however, have several problems. One of the most prevalent EDOF lens is created via wave-front coding using a cubic phase mask. Such an approach generates an Airy beam, which propagates through free space without significant distortion. The resulting PSF, however, does not resemble a point, and images captured with this element are therefore blurry. Computational reconstruction is required to undo the distortion. Another option could be to use a log-asphere lens. For a log-asphere lens, however, different parts of the lens focus at different depths, significantly limiting the focusing efficiency.

In this paper, we design and fabricate an EDOF cylindrical metasurface lens (metalens) using an inverse electromagnetic design methodology. Unlike any existing implementations of diffractive or refractive EDOF lenses, the reported metalens creates a lens-like PSF, without sacrificing any efficiency like many other EDOF lenses. We designed the EDOF lens to have three times the depth of focus of an ordinary lens. The design is validated by fabrication and optical characterization. We found reasonable agreement between both simulation and experiment in terms of the focusing efficiency and the FWHM of the focal spots. The depth of focus of the inverse designed metalens is extended by a factor of 1.5-2. We also did not observe any degradation of the efficiency, as expected from the design process.

4.2 Inverse design and validation

The depth of focus $\Delta f$ of an ordinary lens with diameter $D$ and focal length $f$, for optical wavelength $\lambda$ is given by

$$\Delta f = 4\lambda \frac{f}{D^2}. \quad (4.1)$$

We aim to demonstrate a depth of focus at least three times this value. We specify our Figure of Merit (FOM) as the intensity at eight linearly spaced points along the focal axis which cover an interval of length $3 \times \Delta f$ and centered at the focal spot. We use max-min multi-objective optimization to maximize
We optimize the EDOF lens FOM by adjusting the widths of 150 nanopillars positioned at the center of each lattice cell. The periodicity of the metasurface is kept constant at 443 nm. The thickness of the metasurface is kept constant at 600 nm and we constrain the minimum feature of the metasurface to 100 nm. Lenses with three different focal lengths (66.66 µm, 100 µm, 133.33 µm) are designed.

Figure 4.1: Scanning electron micrograph (SEM) of (top-left) inverse-designed EDOF metalens, (top-right) traditional metalens; (middle-left and middle-right) Zoom-in SEM on inverse-designed EDOF metalens which shows silicon nitride gratings forming cylindrical EDOF metalens; The scale bars correspond to 10 µm (top-left and top-right) and 1 µm (middle-left and middle-right); (bottom) Con-focal microscopy setup used to measure the metalenses.
Figure 4.2: Simulated intensity along the optical axis of the ordinary metalens for 66.66 μm (A), 100 μm (B), and 133.33 μm (C) focal lengths. Experimentally measured filed profile along optical axis for EDOF metalenses with 66.66 μm (D), 100 μm (E), and 133.33 μm (F) focal lengths. (G-O) show cross-section of the beam size in different distance from EDOF metalenses with their Gaussian and double Gaussian fit functions. All numbers which are measured experimentally has ±2 μm error bar.
4.3 Experimental Demonstration

To validate our metasurface design, we fabricated the cylindrical metalenses in silicon nitride. A 600-nm-thick layer of silicon nitride was first deposited on a 500-µm-thick fused-silica substrate using plasma-enhanced chemical vapor deposition (PECVD). The sample was then spin-coated with electron-beam resist (ZEP-520A) and then the metalenses pattern was exposed on it by electron-beam lithography. 8-nm of Au/Pd as a charge dissipation layer was sputtered on the resist prior to electron-beam lithography to prevent pattern distortion due to electrostatic charging. After the lithography step, the charge dissipating layer was removed by type TFA gold etchant and the resist was developed in amyl acetate. A 50 nm layer of aluminum was then evaporated onto the sample. After performing lift-off, the sample was etched using an inductively coupled plasma etcher with a mixture of CHF$_3$ and O$_2$ gases, and then remaining aluminum was removed in AD-10 photoresist developer. To demonstrate the extension of the depth of focus, we fabricated two sets of lenses: one ordinary metalens designed via the usual forward design process and versions developed using our inverse design process. Figs. 4.1(A, B) show the scanning electron micrograph (SEM) of the fabricated EDOF metalens and traditional metalens, respectively. Figs. 4.1(C, D) show zoom-in SEM on inverse-designed EDOF metalens which shows silicon nitride gratings forming cylindrical EDOF metalens. We fabricated three sets of metalenses corresponding to three different focal lengths. The fabricated lenses were measured using a confocal microscopy setup under illumination by 625 nm light-emitting diodes (part number Thorlabs-M625F2), see Fig. 4.1E. Figs. 4.2(A-F) show the simulated and experimentally measured field profiles for the three EDOF metalenses. The intensity profiles along the optical axis are captured using a camera and translating the microscope along the optical axis using an automated translation stage. Figs. 4.2(A-F) also show two-dimensional finite-difference time-domain (FDTD) simulations of the final designed structures match quite well with the experimentally measured focusing behavior. A clear elongation of the focal spot along the optical axis is observed. We fit the intensities near the focal plane...
using a Gaussian function to estimate the full-width-half-maxima (FWHM). Figs. 4.2(G-O) show the Gaussian fit focal spot at the center focal plane, and at two ends of the line along which the light started to be double Gaussian. We identified the depth of focus range when the Gaussian behavior of focal spot started to become double Gaussian. The minimum FWHM for fabricated EDOF metalenses with three different focal lengths (66.66 μm, 100 μm, 133.33 μm) which are shown in Figs. 4.2H, K, N are 1.07 μm, 1.7 μm, 2.32 μm, respectively.

![Figure 4.3: Performance of fabricated ordinary and EDOF metalenses as a function of distance along the optical axis.](image)

By plotting the FWHM as a function of the distance along the optical axis, we estimated the focal length of the metalenses (Figs. 3A-C). We then estimated the focusing efficiency of the lenses along the optical axes. We define the focusing efficiency as the power within a circle with radius of three times the FWHM at the focal plane to the total power incident upon the metalens. The FWHM at the focal
Table 4.1: Comparison between the ordinary and EDOF metalenses for 66.66μm, 100μm, 133.33μm metalenses, respectively.

<table>
<thead>
<tr>
<th>Lens Properties</th>
<th>Ordinary Metalens</th>
<th>EDOF Metalens</th>
</tr>
</thead>
<tbody>
<tr>
<td>FWHM (μm)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulation</td>
<td>0.95</td>
<td>0.97</td>
</tr>
<tr>
<td>Experiment</td>
<td>1.1</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>2.25</td>
</tr>
<tr>
<td></td>
<td>12 (± 0.05)</td>
<td>17 (± 0.05)</td>
</tr>
<tr>
<td></td>
<td>14 (± 0.05)</td>
<td>30 (± 0.05)</td>
</tr>
<tr>
<td></td>
<td>18 (± 0.05)</td>
<td>44 (± 0.05)</td>
</tr>
<tr>
<td></td>
<td>67 (± 0.05)</td>
<td>64.2 (± 0.05)</td>
</tr>
<tr>
<td></td>
<td>102 (± 0.05)</td>
<td>99.7 (± 0.05)</td>
</tr>
<tr>
<td></td>
<td>136 (± 0.05)</td>
<td>129.8 (± 0.05)</td>
</tr>
<tr>
<td>Efficiency (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulation</td>
<td>24.18</td>
<td>24.65</td>
</tr>
<tr>
<td>Experiment</td>
<td>22.26</td>
<td>22.21</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>19.3</td>
</tr>
<tr>
<td></td>
<td>18.6</td>
<td>16.8</td>
</tr>
<tr>
<td></td>
<td>16.8</td>
<td>16.34</td>
</tr>
<tr>
<td></td>
<td>64.2 (± 0.05)</td>
<td>70 (± 2)</td>
</tr>
<tr>
<td></td>
<td>99.7 (± 0.05)</td>
<td>98.2 (± 2)</td>
</tr>
<tr>
<td></td>
<td>129.8 (± 0.05)</td>
<td>134 (± 2)</td>
</tr>
<tr>
<td>Depth of focus (μm)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulation</td>
<td>12 (± 2)</td>
<td>17 (± 2)</td>
</tr>
<tr>
<td>Experiment</td>
<td>16 (± 2)</td>
<td>30 (± 2)</td>
</tr>
<tr>
<td></td>
<td>22 (± 2)</td>
<td>44 (± 2)</td>
</tr>
<tr>
<td>Focal Length (μm)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulation</td>
<td>67 (± 2)</td>
<td>64.2 (± 2)</td>
</tr>
<tr>
<td>Experiment</td>
<td>102 (± 2)</td>
<td>99.7 (± 2)</td>
</tr>
<tr>
<td></td>
<td>136 (± 2)</td>
<td>129.8 (± 2)</td>
</tr>
</tbody>
</table>

plane is calculated as the minimum FWHM from the Figs. 4.3(A-C). We plot the focusing efficiency of the metalenses along the optical axis (Figs. 4.3(D-F)). We expect the focusing efficiency to remain the same along the depth of the focus, and then drop off as we go away from the depth of focus. Clearly, for the EDOF metalens, the efficiency remains constant over a longer depth as expected. Table 4.1 summarizes the performance of all the metalenses, in terms of FWHM, focal length, efficiency, and depth of focus. We find a reasonable agreement between the simulation and experimental results. Additionally, we clearly observed an extended depth of focus in the inverse-designed metasurfaces compared to the ordinary metalenses. We also observe no significant efficiency degradation between the ordinary metalenses and the EDOF metalenses.

4.4 Discussion

We demonstrated for the first-time inverse designed EDOF cylindrical metalenses. While several theoretical designs exist for cylindrical EDOF lenses, to the best of our knowledge no experimen-
tally demonstrated EDOF cylindrical lenses have been reported before. While we extended the depth by a factor of 2-3, the depth of focus can be further extended albeit at the cost of reduced efficiency. In this work, we focused on 1D cylindrical lenses for the simplicity of design, and the next step will be to extend this concept to 2D lenses and demonstrate imaging over a broader optical bandwidth, than what is possible using a traditional metalens. Going beyond extended depth of focus, the inverse design techniques can be used to engineer PSF of the optics with far-reaching impact on computational imaging and microscopy.
In this chapter, we extend the framework from Chap. 3 to three-dimensional large-area lenses and show experimental results of achromatic lenses with large diameter. The main contributions of our

work are employing inverse design for achromatic applications and fabricating large area metasurfaces with two techniques, e-beam lithography and deep-UV lithography.

This chapter follows previous work in metalens for applications for visible wavelengths. Some work has demonstrated achromatic metalens for visible wavelengths, however due to the design limit, the metalens is quite small. The largest achromatic metalens prior to our work has a diameter of no more than 50 μm. Our current work is able to extend the diameter to millimeter and centimeter scales for achromatic lenses. We achieve this performance for two reasons: 1) our inverse design framework easily extends to large diameters, and 2) for the centimeter scale, we use a newly-demonstrated KrF Deep-UV lithography technique to fabricate very large metalenses (diameters of 1-2 cm) at a low cost; however, this technique has only been applied to chromatic lenses as of yet.

5.1 Theory

The inverse design framework for metasurfaces is extended to three dimension is this chapter. We first introduce extensions to the fast approximate solver and then introduce improvements regarding the optimization problem: the inverse design.

5.1.1 Fast approximate solver

The approximate solver from Chap. 3 can be directly extended from two dimensions to three dimensions by computing the convolution between LPA sources and the Green’s function (Eq. 3.2) with two-dimensional Cartesian coordinates.

Cylindrical symmetry in the parameter function In order to speed up the optimization process, we use rotational symmetry for the Green’s function convolution in polar coordinates. Effectively, the two-dimensional integral becomes a one-dimensional integral. When computing the electric field
at the focal point, Eq. 3.2, becomes

$$E_z(x_{\text{focal}}) = 2\pi \int_{\text{surface}} G(x_{\text{focal}}, r) \text{source(\text{param}(r)) } r dr; \quad (5.1)$$

where $x_{\text{focal}}(0, 0, \text{focal})$ is the target focal spot, and \text{param} is the parameter function of the surface on the radial axis. Since the focal point is on the axis of symmetry of the surface, both the source and the Green’s function are only functions of the radial coordinate. Therefore, the angle integral simply gives a factor of $2\pi$. The integral, which would be a two-dimensional integral in Cartesian coordinates, becomes a one-dimensional integral thanks to the symmetry of our setup, which dramatically speeds up the computation of the objective function, since the complexity becomes $o(\sqrt{n})$, where $n$ is the number of unit cells to optimize. The drawback of the cylindrical symmetric approximate solver is that it can only solve for normal incidence on the axis of symmetry of the metasurface because the Green’s function is no longer cylindrical symmetric otherwise.

For non-normal incidence or evaluation of points outside the focal axis, it is possible to keep a cylindrical symmetric parameter function and compute the Green’s function integral $G(r)$ off-line, as in

$$E_z(x_{\text{focal}}) = \int_{\text{surface}} G(r) \text{source(\text{param}(r)) } r dr; \quad (5.2)$$

where $G(r) = \int G(x_{\text{focal}}, r, \theta) d\theta$ and the integral is on the support $[0, 2\pi]$.

**Choice of polarization:** Right-circular polarized (RCP) conversion to Left-circular polarized (LCP) Originally, we wanted to capitalize on the geometrical Pancharatman-Berry (PB) phase.\textsuperscript{104,105} The PB phase is a geometrical phase added to the field as one rotates the unit cell. It only exist for the RCP conversion to LCP. For all optimization we used the basis RCP, LCP to decompose the field instead of the traditional linear polarization basis. However, the gain from PB phase did not outweigh the loss from limiting the design parameters to the circumcircle of the unit cell’s square.
We therefore fixed the angle and allowed the parameters to go up to fabrication constraints.

**Polarization-insensitive design**  
Thanks to the RCP/LCP decomposition, the only geometry needs to have mirror symmetry to achieve polarization-insensitivity. If you consider a solution of Maxwell’s equations corresponding to a “RCP to LCP conversion”, then a mirror flip will not change the geometry, but it will change the solution to an “LCP to RCP conversion”. Therefore, since our geometry is rotational symmetric (which includes mirror symmetric), our designs are polarization-insensitive.

**Surrogate model**  
We use Chebyshev interpolation as a surrogate model. In the application with 2mm diameter, the unit cell has 2 parameters which are evaluated at 15\(^2 = 225\) Chebyshev coordinates. In the application with 1 cm diameters, the unit cell has one parameter which is evaluated at 20 Chebyshev points.

5.1.2 **Inverse design**

**Bigger resolution than the unit cell’s period**  
Increasing the number of degrees of freedom usually results in better design.\(^4\) We increase the effective number of degrees of freedom by increasing the sampling resolution of the parameter function along the radial axis finer than the unit cell period. As showed in Eq. 5.1, we use a cylindrical symmetry to speed up the approximate solver. However, we still need to evaluate the function to find the parameters of the unit cells which are on a square mesh, truncated in order to make it a circular surface with the desired diameter. There are two extreme radial axes which require different sampling of the parameter function: 1) sampling at every unit cell period if we go along a radial axis which passes through the middles of two of the unit cells’ sides, and 2) sampling at every \(\sqrt{2}\) times the unit cell period if we go along a radial axis which passes through all the unit cell diagonals. We also need to sample the radial axis at every angle between those two case
which covers the interval \([0, 45] \) degrees. This feature stems from the fact that we now optimize for a three-dimensional lens versus a two-dimensional lens as in Chap. 3. In Chap. 3, we needed only one degree of freedom per unit cell period on the metasurface. In this chapter, we can add degrees of freedom on the radial axis by having a resolution of the parameter function finer than the unit cell’s period. In practice, we use between 3 and 10 points per unit cell period. Then, when evaluating the parameter function on a grid, we do a linear interpolation of this finer function. Increasing the resolution has physical meaning but also increases the number of degrees of freedom for the optimization, leading to better local optima in our experience.

**Optimal objective wavelengths**  
The best choice of constraints for an achromatic application is to space them linearly in the frequency domain.† Since we are solving the wave equation in the frequency domain. For a metalens with chromatic aberration, the focal shift is proportional to the frequency \((dF = K dw)\). When designing a multi-wavelength achromatic metalens, we want to minimize the interference between the main focal point and the spurious focal point. Therefore, we want to push their focal distances as far as possible. The focal distance is determined by the frequency difference between neighboring focusing frequencies. For a compromise to cover a bandwidth best, we need to constrain equally-spaced frequency instead of equally-spaced wavelength. Following this line of thought, we used constraints at 490 nm, 520 nm, 540 nm, 570 nm, 610 nm, and 650 nm for our achromatic application in the visible constraining on six wavelengths.

**Approximation: optimizing the zeroth order coefficient in the presence of diffractive orders.** When the unit cell is large enough to diffract higher orders as in Sec. 5.3.3, we approximate the intensity at the focal spot using only the zeroth order Fourier mode. This approximation

---

†I would like to thank my father Didier Pestourie—computer science engineer—for noticing this fact as he was reading Ref. 3. Thanks to his outstanding wit, he noticed that the intensity lines from Fig. 3.5 have the pattern of hyperbolas, and he advised me to space the constraints linearly in inverse wavelengths.
is valid when the numerical aperture is smaller than \( \cos(\arcsin(\lambda/D)) \); we prove this result in this paragraph.

We show in which regime this approximation is valid by finding the minimum numerical aperture (NA) of a lens, for the first diffractive order to have an effect on the intensity at the focal spot.

For normal incidence, the grating equation is

\[
D(\sin \delta_m) = m\lambda,
\]

where \( D \) is the diameter of the lens, \( \lambda \) is the wavelength of the incident light, \( m \) is the order of diffraction, and \( \delta_m \) is the angle of propagation of the diffractive order.

By definition, the numerical aperture for propagation in vacuum is

\[
NA = \sin \arctan \frac{D}{2f}
\]

where \( f \) is the focal length of the lens.

Therefore, the minimum numerical aperture needed for the first diffractive order to have an impact on the focal spot is

\[
NA_{\text{min}} = \sin \frac{\pi}{2} \delta_1 = \cos \delta_1 = \cos \arcsin \frac{\lambda}{D}.
\]

In Table 5.1, we show the minimum numerical aperture needed for our approximation to fail. Since we are interested in an application with numerical aperture of \( NA = 0.2 \), we are in the range where our approximation is valid for the design with three wavelengths (488, 532, and 658 nm).

**Examples of design parameter functions** This paragraph contains a high level overview of the optimization problem and examples of resulting designs. Our main message here is that the result-
### Table 5.1: Minimum NA needed before a diffractive order contributes to the focal spot.

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>NA&lt;sub&gt;min&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>488 nm</td>
<td>0.82</td>
</tr>
<tr>
<td>532 nm</td>
<td>0.8</td>
</tr>
<tr>
<td>658 nm</td>
<td>0.63</td>
</tr>
</tbody>
</table>

**Figure 5.1:** (A) 2 mm diameter lens unit cell with TiO<sub>2</sub> nanofin on top of a silica substrate with period 400 nm. Each unit cell has two parameters—the lengths and a width of the nanofin—and is fabricated using e-beam lithography. (B) 1 cm diameter lens unit cell with silica nanopost on top of a silica substrate with period 850 nm. Each unit cell has one parameter—the radius of the post—and is fabricated using UV lithography.

Ining designs are exploiting all the degrees of freedom in non-trivial and oscillatory manner, therefore showing that inverse design framework works. The subsequent sections will present the fabrication, the performance, and the characterization of the resulting lens applications in more detail.

We considered two types of unit cells for a diameter of 2 mm and a diameter of 1 cm. The smaller metasurface consists of unit cells with TiO<sub>2</sub> nanofin on top of a silica substrate with period 400 nm (Fig. 5.t(A)). Each unit cell has two parameters—length and width—and is fabricated using e-beam lithography. The larger metasurface consists of unit cells with silica nanopost on top of a silica substrate with period 850 nm (Fig. 5.t(B)). Each unit cell has one parameter—the radius of the post—and is fabricated using deep UV lithography (newly-tested for large diameter metasurfaces in Ref. 102).
We solved a multiwavelength optimization problem for intensity maximization

$$\max_{\text{parameters}} \left[ \min_{l \in L} I_{\text{target}}(\text{parameters}, l) \right] .$$

where $I$ is the intensity, $L$ is the set of wavelengths containing red, green, and blue (we also considered the set \{490, 520, 540, 570, 610, 650\} nm). Depending on the application, the target is defined as the focal spot for numerical apertures in \{0.2, 0.25, 0.3, 0.7\}, parameters are sampling points of the parameters function in the radial direction (three points per unit cell period).

The optimized parameters function for a 2 mm diameter lens (Fig. 5.2(A)), and a 1 cm diameter lens (Fig. 5.2(B)) both achromatic for red, green, and blue are very oscillatory and non-trivial. In order to express the full usage of the two dimensions of the 2 mm diameter lens unit cell, the aspect ratio and the area are plotted instead of the length and the width. In the 1 cm diameter application, insets show zooms of the parameter function at the center and for a larger radius, in both cases the
parameter is oscillatory while remaining non-trivial. This non-triviality and the use of all degrees of freedom available are characteristics of inverse design. The locally optimal solutions could not be foreseen in advance.

5.2 Methods

The fabrication and measurement were performed with my collaborator, Dr. Zhaoyi Li (Harvard).

5.2.1 Measurement method

The light source used in the metalens focal shift measurement is a SuperK continuum fiber-coupled laser as shown in Fig. 5.3. The incident beam from the laser source is collimated by a fiber-coupled collimator (RC12APC-01, Thorlab). The collimated beam is incident on the metalens from the backside and focuses on the other side. The focal spot is captured by a microscope setup that consists of an objective lens (Olympus MPlan Apo N) and a tube lens (TTL180-A, Thorlab). The numerical aperture (NA) of the objective lens is 0.95 that is larger than the NA of the metalens. The intensity distribution of the focal spot is recorded by a CMOS camera (DCC1545M, Thorlab). The objective lens, the tube lens, and the CCD camera are mounted on the transitional stage that can move in three dimensions. During the measurement, the microscope system scans in z-direction to record the focusing intensity profile along the optical axis of the metalens. To characterize the Strehl ratio of the focal spot of the metalens, we used diode laser with narrow linewidth as the light source. The linewidth of the diode laser beam is less than 1 nm. The output beam from the diode laser is coupled into a fiber using a fiber coupler and couples out as a collimated laser beam through a fiber coupled collimator. The rest of the measurement setup is identical as described above.

The setup using diode laser can be found in Appendix B.
5.2.2 Fabrication method

The metalens was fabricated following the procedures as described below:

1. A layer of 600 nm-thick positive e-beam resist (EBR, Zeon Chemicals, ZEP-520A) was spun onto a fused silica substrate, followed by a second layer of conductive polymer (Showa Denko, ESPACER 300) in order to reduce the charging effects during the electron beam lithography (EBL).

2. The pattern was transferred onto the resist-coated substrate by using EBL (Elinonix, ELS-F125) with exposure current of 1 nA. The exposed sample was then developed in o-xylene.

3. Around 210 nm-thick amorphous TiO$_2$ was uniformly deposited onto the developed sample using atomic layer deposition (ALD, Cambridge Nanotech, Savannah) method. The exposed trenches in the EBR were filled with TiO$_2$.

4. We removed the residue TiO$_2$ deposition layer on top of the unexposed sample area using reactive ion etching (RIE, Oxford Instruments, PlasmaPro 100 Cobra 300). The etching gas is a mixture of CHF$_3$, O$_2$, and Ar gas.
Table 5.2: Specifications of the achromatic lens for red, green, and blue: [488, 532, 658] nm with 2 mm diameter.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelengths</td>
<td>488, 532, and 658</td>
</tr>
<tr>
<td>Unit cell period</td>
<td>400 nm</td>
</tr>
<tr>
<td>Number of unit cells</td>
<td>~ 25 millions</td>
</tr>
<tr>
<td>Diameter</td>
<td>2 mm</td>
</tr>
<tr>
<td>Focal length</td>
<td>4.88-9.7 mm</td>
</tr>
<tr>
<td>Numerical aperture</td>
<td>0.1-0.2</td>
</tr>
<tr>
<td>Nanofin width range</td>
<td>50 - 350 nm</td>
</tr>
<tr>
<td>Nanofin length range</td>
<td>50 - 350 nm</td>
</tr>
<tr>
<td>Optimization time</td>
<td>~1h on a laptop</td>
</tr>
</tbody>
</table>

5. Finally, we remove the EBR by soaking the sample in the RG remover (MicroChem Corporation) bath for 24 hrs at room temperature.

The Strehl ratio is computed from fitting an Airy function. 107

5.3 Experimental results and discussion

Three achromatic lens applications were designed, fabricated, and measured. The unit cell simulations needed to train our offline solver were performed by my collaborator, Dr. Zhaoyi Li at Harvard University, using Lumerical. 108

5.3.1 Achromatic lens for RGB with 2 mm diameter

Our first application is an achromatic lens for red, green and blue (RGB): [488, 532, and 658] nm with a diameter of 2 mm. The process of fabrication uses electron-beam lithography. The table of specification for this application is in Table 5.2.

We validated our approximate solver against brute force simulation with the 2 mm unit cell on a single wavelength design for 530 nm and a 10 μm diameter. The FDTD simulation in Fig. 5.4(A-B) shows clear focusing both on the XZ cut Fig. 5.4(A) and along the focal axis Fig. 5.4(B). Because of
symmetry, the approximate solver is only fit to compute the intensity along the focal axis Fig. 5.4(C).
We see very good agreement between the two solvers in the far field, and as expected, the solvers differ significantly close to the metasurface (at a distance less than 5 μm away from the metasurface). We optimized for a lens with 2 mm diameter (with a refinement of 3 per wavelength) and for red, green, and blue (Fig. 5.4(D)). The intensity along the focal axis shows clear focusing at the focal spot for the three desired wavelengths [488, 532, and 558] nm. The numerical aperture of this lens is 0.2; a diameter of 1 cm corresponds to a focal length of 2.5 cm.

We measure the focusing intensity distribution of the metalens in the XZ cross section at three designed wavelengths: 488 nm, 532 nm and 658 nm under different polarization configurations. (x-direction is along lens radius and z-direction is along the optical axis). The incident beam is coupled from a SuperK continuum fiber laser and has a bandwidth of ∼5 nm. In Fig. 5.5(left), the incident beam is unpolarized and focused by the metalens. The output beam intensity profile is captured by the CCD camera through an optical relay system without passing through any polarizer. In Fig. 5.5(middle), the incident beam is selected to be left-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the right-handed circularly-polarized component. In Fig. 5.5(right), the incident beam is selected to be right-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the left-handed circularly-polarized component.

We characterize the lens by measuring the intensity of the field along a line parallel to the metasurface and going through the focal point (Fig. 5.6(A)). Then, we fit an Airy function to compute the Strehl ratio. A Strehl ratio bigger than 0.8 signifies that the lens is diffraction-limited, which it is for the three wavelengths [488, 532, and 658] nm when converting from RCP to LCP. In Fig. 5.6(B), we show the intensity of the field measured on a cut parallel to the metasurface and going through the focal point. We see an error pattern compared to an Airy disk with a \( C^4 \) symmetry. It comes from the squared lattice which breaks the \( C^\infty \) symmetry of the parameter function. The error from the local
Figure 5.4: We validated our approximate solver against brute force simulation with the 2 mm unit cell on a single wavelength design for 530 nm and a 10 µm diameter. The FDTD simulation (A-B) shows clear focusing both on the xz-cut (A) and along the focal axis (B). Because of symmetry, the approximate solver is only fit to compute the intensity along the focal axis (C). The two solvers match very well in the far field, and as expected, differ significantly close to the metasurface (at a distance less than 5 µm away from the metasurface). (D) We optimized for a lens with 2 mm diameter (with a refinement of 3 per wavelength) and for red, green, and blue. The intensity along the focal axis shows clear focusing at the focal spot for the three desired wavelengths [488, 532, and 558] nm. The numerical aperture of this lens is 0.2 (focal length of 2.5 cm with diameter of 1 cm).

periodic approximation has the same symmetry as the lattice.

We characterize the lens by measuring the intensity of the field along a line parallel to the metasurface and going through the focal point (Fig. 5.7(A)). Then we fit an Airy function to compute the Strehl ratio. A Strehl ratio bigger than 0.8 signifies that the lens is diffraction limited, which it is for the
Figure 5.5: Measurement results of focusing intensity distribution of the metalens in the XZ cross section at three designed wavelengths: 488 nm, 532 nm and 658 nm under different polarization configurations. (\(x\)-direction is along lens radius and \(z\)-direction is along the optical axis). The incident beam is coupled from a SuperK continuum fiber laser and has a bandwidth of \(\sim\) 5 nm. (Left) The incident beam is unpolarized and focused by the metalens. The output beam intensity profile is captured by the CCD camera through an optical relay system without passing through any polarizer. (Middle) The incident beam is selected to be left-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the right-handed circularly-polarized component. (Right) The incident beam is selected to be right-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the left-handed circularly-polarized component.

The results show the intensity of the field measured on a cut parallel to the metasurface and going through the focal point. We see an error pattern compared to an Airy disk with a \(C_4^v\) symmetry. It comes from the squared lattice which breaks the \(C_\infty^v\) symmetry of the parameter function. The error from the local periodic approximation has the same symmetry as the lattice. Apart from 658 nm wavelength, which shows more side lobes, the results are very similar to the conversion from LCP to RCP.

Optical microscope images of the fabricated metalens show stitching error which might explain the asymmetry in the error pattern. In Fig. 5.8(Top panel), we present optical microscope images at 50\(\times\) magnification. In Fig. 5.8(Bottom panel), we present optical microscope images at 100\(\times\) magnification. As can be seen from the images, the fabricated metalens suffers from minor stitching errors. The
Figure 5.6: (A) We characterize the lens by measuring the intensity of the field along a line parallel to the metasurface and going through the focal point. Then we fit an Airy function to compute the Strehl ratio. A Strehl ratio bigger than 0.8 signifies that the lens is diffraction-limited, which it is for the three wavelengths [488, 532, and 658] nm when converting from RCP to LCP. (B) We show the intensity of the field measured on a cut parallel to the metasurface and going through the focal point. We see an error pattern compared to an Airy disk with a \( C_4 \) symmetry. It comes from the squared lattice which breaks the \( C_4 \) symmetry of the parameter function. The error from the local periodic approximation has the same symmetry as the lattice.

stitching errors occur at the boundaries of e-beam writing fields (area of 250 μm × 250 μm) during e-beam lithography process. The stitching error line is across the outer part of the metalens and will introduce wavefront errors. The rest part of the metalens is intact. We annotated the stitching error boundaries with red arrows.

5.3.2 ACHROMATIC LENS FOR SIX WAVELENGTH WITH 2 MM DIAMETER

Our second application is an achromatic lens for [490, 520, 540, 570, 610, and 650] nm, with a diameter of 2 mm. As explained in the theory section, the wavelengths are chosen to be linearly-spaced in the frequency domain. The process of fabrication uses electron-beam lithography. The table of specification for this application is in Table. 5.3.
Figure 5.7: (A) We characterize the lens by measuring the intensity of the field along a line parallel to the metasurface and going through the focal point. Then we fit an Airy function to compute the Strehl ratio. A Strehl ratio bigger than 0.8 signifies that the lens is diffraction limited, which it is for the three wavelengths [488, 532, and 658] nm for unpolarized light. (B) We show the intensity of the field measured on a cut parallel to the metasurface and going through the focal point. We see an error pattern compared to an Airy disk with a $C_4$ symmetry. It comes from the squared lattice which breaks the $C_4$ symmetry of the parameter function. The error from the local periodic approximation has the same symmetry as the lattice. Apart from 658 nm wavelength which shows more side lobes, the results are very similar to the conversion from LCP to RCP.

Table 5.3: Specifications of the achromatic lens for six wavelength across visible range with 2 mm diameter.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelengths</td>
<td>490, 520, 540, 570, 610, and 650 nm</td>
</tr>
<tr>
<td>Unit cell period</td>
<td>400 nm</td>
</tr>
<tr>
<td>Number of unit cells</td>
<td>~25 millions</td>
</tr>
<tr>
<td>Diameter</td>
<td>2 mm</td>
</tr>
<tr>
<td>Focal length</td>
<td>4.88-9.7 mm</td>
</tr>
<tr>
<td>Numerical aperture</td>
<td>0.1-0.2</td>
</tr>
<tr>
<td>Nanofin width range</td>
<td>50 - 350 nm</td>
</tr>
<tr>
<td>Nanofin length range</td>
<td>50 - 350 nm</td>
</tr>
<tr>
<td>Optimization time</td>
<td>~4h on a laptop</td>
</tr>
</tbody>
</table>
After optimization, the designed intensity along the focal axis shows clear focusing at the focal spot for the six desired wavelengths, which are linearly-spaced in the frequency domain to achieve best achromatic properties in the visible: [490, 520, 540, 570, 610, and 650] nm. In Fig. 5.9(A), the numerical aperture of this lens is 0.1; a diameter of 2 mm corresponds to a focal length of about 1 cm. The lens shows multiple spurious foci along the focal axis. Although the focal spot is clearly demarcated from the spurious foci, the spurious intensities can be bigger than the intensity at the desired focal spot. In Fig. 5.9(B), the numerical aperture of this lens is 0.2; a diameter of 2 mm corresponds to a focal length of about 5 mm. This lens show similar features as the NA = 0.1 case. Some measurements of the focal length on a bigger support can be found in Appendix. B.
We measure the focusing intensity distribution of the metalens of $\text{NA} = 0.2$ in the XZ cross section at six designed wavelengths: 490 nm, 520 nm, 540 nm, 570 nm, 610 nm, and 650 nm under different polarization configuration. ($x$-direction is along lens radius and $z$-direction is along the optical axis.) The incident beam is coupled from a supercontinuum (SuperK) fiber laser and have a bandwidth of $\sim$5 nm. In Fig. 5.10(a), the incident beam is unpolarized and focused by the metalens. The output beam intensity profile is captured by the CMOS camera through an optical relay system without passing through any polarizer. In Fig. 5.10(b), the incident beam is selected to be left-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the right-handed circularly-polarized component. In Fig. 5.10(c), the incident beam is selected to be right-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the left-handed circularly-polarized component.

We measure the intensity profiles at the focal planes of the 6-wavelength achromatic metalens for design wavelengths of 490 nm, 520 nm, 540 nm, 570 nm, 610 nm, and 660 nm, respectively, in Fig. 5.11. The incident light is from a supercontinuum (SuperK) laser and is unpolarized. The bandwidth of the
Figure 5.10: Measurement results of focusing intensity distribution of the metalens of NA = 0.2 in the xz-cross section at six designed wavelengths: 490 nm, 520 nm, 540 nm, 570 nm, 610 nm, and 650 nm under different polarization configuration. (x-direction is along lens radius and z-direction is along the optical axis). The incident beam is coupled from a supercontinuum (SuperK) fiber laser and have a bandwidth of ~5 nm. (a) The incident beam is unpolarized and focused by the metalens. The output beam intensity profile is captured by the CMOS camera through an optical relay system without passing through any polarizer. (b) The incident beam is selected to be left-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the right-handed circularly-polarized component. (c) The incident beam is selected to be right-handed circularly-polarized, and the output beam passes through a quarter waveplate followed by a linear polarizer to select the left-handed circularly-polarized component.

light source is around 5 nm. The intensity profiles of the focal spots are captured by a CMOS camera through an optical relay system (i.e. microscope setup) without passing through any polarizer.

Since we do not have access to the measurements using diode laser, we need to characterize the lens differently from the previous application. We measure the intensity profile of the metalens in line cut through the center of focal spots at design wavelengths of 490 nm, 520 nm, 540 nm, 570 nm, 610 nm, and 650 nm and compare it to the ideal Airy disk at the same wavelength. In Fig. 5.12, the solid lines are measurement results and the dashed lines are intensity profiles of ideal Airy Disks at design wavelengths for comparison. The discrepancy shows up at the tails of the focusing peaks. The reason is that the incident light source has a finite bandwidth, and the focal spot shifts away from the design focal length when wavelength deviates from the design wavelengths.
Figure 5.11: Measured intensity profiles at the focal planes of the 6-wavelength achromatic metalens for design wavelengths of 490 nm, 520 nm, 540 nm, 570 nm, 610 nm, and 660 nm, respectively. The incident light is from a supercontinuum (SuperK) laser and is unpolarized. The bandwidth of the light source is around 5 nm. The intensity profiles of the focal spots are captured by a CMOS camera through an optical relay system (i.e. microscope setup) without passing through any polarizer.

In Fig. 5.13, we show the measured frequency width at half maximum (FWHM) of the normalized focusing intensity profiles at design wavelengths versus the theoretical FWHM of the normalized focusing intensity profiles of ideal Airy disks at corresponding wavelengths. They agree with each other well at long wavelengths of 610 nm and 650 nm and show differences at short wavelengths of 490 nm, 510 nm, 540 nm, and 570 nm. The maximum difference is \( \sim 19\% \) of the theoretical value at the wavelength of 490 nm.

5.3.3 Achromatic lens for RGB with 1 cm diameter

Our last application is an achromatic lens for red, green, and blue (RGB): [488, 532, 658] nm with a diameter of 1 cm. Going to 1 cm diameter is a remarkable breakthrough. The table of specification for this application is in Table 5.4.
Figure 5.12: Measured intensity profile of the metalens in line cut through the center of focal spots at design wavelengths of 490 nm, 520 nm, 540 nm, 570 nm, 610 nm, and 650 nm. The solid lines are measurement results and the dashed lines are intensity profiles of ideal Airy Disks at design wavelengths for comparison. The discrepancy shows up at the tails of the focusing peaks. The reason is that the incident light source has a finite bandwidth, and the focal spot shifts away from the design focal length when wavelength deviates from the design wavelengths.

Table 5.4: Specifications of the achromatic lens for red, green, and blue with 1 cm diameter.

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelengths</td>
<td>488, 532, 658 nm</td>
</tr>
<tr>
<td>Unit cell period</td>
<td>850 nm</td>
</tr>
<tr>
<td>Number of unit cells</td>
<td>~ 100 millions</td>
</tr>
<tr>
<td>Diameter</td>
<td>1.00011 cm</td>
</tr>
<tr>
<td>Focal length</td>
<td>1.5-2.5 cm</td>
</tr>
<tr>
<td>Numerical aperture</td>
<td>0.2-0.3</td>
</tr>
<tr>
<td>Pillar radius range</td>
<td>250 - 600 nm</td>
</tr>
<tr>
<td>Optimization time</td>
<td>~ 1h on a laptop</td>
</tr>
</tbody>
</table>

The unit cell of the one centimeter lens has a period of 850 nm, which is bigger than the design wavelengths [488, 532, 658] nm and will diffract non-zero orders. It consists of a cylindrical pole and has the radius of the cylinder as its single parameter. Because of the fabrication process, the pillar is not a perfect rectangle as seen in Fig. 5.14(A); rather, the side is somewhat slanted at the top. We take
Figure 5.13: The measured FWHM of the normalized focusing intensity profiles at design wavelengths versus the theoretical FWHM of the normalized focusing intensity profiles of ideal Airy disks at corresponding wavelengths. They agree with each other well at long wavelengths of 610 nm and 650 nm, and show differences at short wavelengths of 490 nm, 510 nm, 540 nm, and 570 nm. The maximum difference is 19% of the theoretical value at the wavelength of 490 nm.

...this effect into account when computing the local field. We can see a scanning electron microscope image (Fig. 5.14(B)) of the a lens consisting of this unit cell showing this feature. We validated the approximate solver on a structure with diameter 20 μm, which is the biggest we can do in a reasonable amount of time using brute force FDTD (Fig. 5.14(C)).

After optimization, the designed intensity along the focal axis shows clear focusing at the focal spot for the three desired wavelengths [488, 532, and 558] nm (Fig. 5.15). The numerical aperture of this lens is 0.2; a diameter of 1 cm corresponds to a focal length of 2.5 cm. The lens shows spurious foci at around 2.3 cm for the 532 nm wavelength, and 2.7 cm for the 488 nm wavelength; their intensity is...
Figure 5.14: UV-litography makes the wall of the nanopost slanted as showed in (A) the schematics of the unit cell and in (B) a scanning electron microscope image of a fabricated lens, in (C) the geometry of the FDTD simulations that validated our approximate solver.

less than half of the intensity at the desired focal spot. The focal line was also computed on a larger support along the focal axis and showed similar features. This work is still ongoing, and devices have been optimized also for $\text{NA} = 0.25$, and $\text{NA} = 0.3$. As the numerical aperture increases, the intensity of the side peak decreases relative to the focal spot.

5.4 Conclusion

This work is the first demonstration of achromatic metasurfaces with as large diameters (2 mm and 1 cm) and using inverse design. Further figures about measurement setup and field computation can be found in Appendix B.
Figure 5.15: The intensity along the focal axis of a 1 cm diameter lens achromatic for [488, 532, and 558] nm with a numerical aperture of 0.2 shows clear focusing, but also presents spurious foci.

Validation of approximate solver in 3D The quasi-perfect agreement between simulations with our approximate solver and experiments are strong validations of our solver. In Chap. 6 and 7, we introduce methods to train a surrogate model for the unit cell. To validate our approximate solver relying on our surrogate model, we will use our approximate solver relying on an online solver (see Chap. 2) instead of a brute force solver.
It does not say in the Bible that all laws of nature are expressible linearly!

Enrico Fermi

6

Neural network surrogate for inverse design of metasurfaces.

In this chapter,* we present results from our optimization framework when using a predictive neural network (PNN)¹⁰⁹ as offline Maxwell’s equations solver. In contrast, Chebyshev interpolation which requires training points exponentially in the number of variable, neural networks are reported to over-

*Manuscript in preparation.
come the curse of dimensionality.\textsuperscript{10,11}

An et al. have presented a predictive neural network that correctly estimates the local field of a unit cell that has 5 parameters.\textsuperscript{109} In collaboration with them, we incorporated their neural network into our optimization framework (Chap. 3) for three-dimensional metasurfaces (Chap. 5).

6.1 Neural network electromagnetic surrogate model and the unit cell

In this section, we summarize important results about the surrogate model presented in Ref. 109 that we will feed into our large-scale optimization framework in the next section.

6.1.1 Unit cell

The unit cell is showed in Fig. 6.1. Fig. 6.1(A) illustrates the general schematic of the all-dielectric structures under consideration. They all consist of a dielectric meta-atom (upper layer) in the shape of the letter H and a dielectric substrate (bottom layer). During the modeling process, the meta-atoms are arranged in rectangular lattices. Random parameter combinations including the gap, thickness, radius, and permittivity of meta-atoms were generated in the multi-paradigm numerical computing tool MATLAB and then transferred to commercial software package, CST Microwave Studio,\textsuperscript{112} for full-wave simulations.\textsuperscript{109} The period of the unit cell is 3 µm. The supports of the parameters are in Table 6.1.
Table 6.1: Support of parameters in the H-shaped unit cell

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lx</td>
<td>[0.5, 2.7] µm</td>
</tr>
<tr>
<td>Ly</td>
<td>[0.5, 2.7] µm</td>
</tr>
<tr>
<td>Lx1</td>
<td>[0.5, 2.5/3] µm</td>
</tr>
<tr>
<td>Ly1</td>
<td>[0.5, 2.5/3] µm</td>
</tr>
<tr>
<td>Index</td>
<td>[3.5, 5] µm</td>
</tr>
<tr>
<td>Thickness</td>
<td>[0.5, 1] µm</td>
</tr>
</tbody>
</table>

6.1.2 Predictive Neural Network (PNN)

The PNN consists of a bi-linear tensor layer, followed by four consecutive fully-connected layers. The output of the bi-linear tensor layer is given by

\[
\text{Output} = f(e^T W^{[1:k]}_1 e + e^T W^{[1:k]}_2 (e \odot e) + V \begin{bmatrix} e \\ (e \odot e) \end{bmatrix} + b),
\] (6.1)

where \( f \) is the activation function, \( b \) is the bias, and \( e \) is the vector of parameters. \( W^{[1:k]} \) is a tensor of weights to learn of dimension \( n \times n \times k \), where \( n \) is the number of parameters and \( k \) is an hyper-parameter of the model. This layer implements a Neural Tensor Network (NTN)\(^{113}\) instead of a fully-connected first layer. Compared to standard neural networks where the entity vectors are simply concatenated, the main advantage of this NTN is that it can relate the two inputs multiplicatively instead of only implicitly through nonlinearity,\(^{109}\) such as getting the permittivity from squaring the refractive index, or learning areas and volumes. A ReLU activation function is applied to the output tensor of each layer except for the last one, which is fed to a tanh activation function. The real and imaginary parts of the complex transmission coefficient are predicted using two independent networks that share the same network architecture.\(^{109}\) The PNN is implemented using the Tensorflow\(^{114}\) framework.

Fig. 6.1(B) shows the architecture of the neural network that is used as a surrogate model for the so-
A B

Figure 6.1: (A) Top-view figure showing the parameters of the H-shaped meta-atom structure/unit cell. The unit cell has 5 parameters: Lx, Ly, Lx1, Ly1, and the permittivity of the structure. The predictive neural network (PNN) consists of a bi-linear tensor layer, followed by four consecutive fully-connected layers. A ReLU activation function is applied to the output tensor of each layer except for the last one, which is fed to a tanh activation function. The real and imaginary parts of the complex transmission coefficient are predicted using two independent networks that share the same network architecture (B). Details, including four fully-connected hidden layers and five weight arrays, are given in the figure. Courtesy of Sensong An.

olutions to Maxwell’s equation with this unit cell. The neural network was trained with thousands of training inputs.

6.2 Optimization and validation

In this section, we show that optimizing using our optimization framework from Chap. 5 based on the neural network surrogate model, also optimizes the solutions of Maxwell’s equations (as desired).

6.2.1 Optimization

As discussed in Chap. 5, we use the cylindrical symmetric solver to compute the approximate field and its intensity at the focal spot. Based on the network of An et al., we evaluated the local field and
implemented the gradients within the Tensorflow\textsuperscript{144} framework. This local field function and gradients take \(\sim\)1ms to evaluate and are then hooked into our computational framework. We optimized a structure with period 3 \(\mu\)m and diameter 1 mm (\(\sim\)85,5k unit cells) for an achromatic lens over 7 wavelengths from 5 \(\mu\)m to 10 \(\mu\)m. The resolution of the parameter function is 1 \(\mu\)m (a third of the period as discussed in Chap. 5). The optimization took 11 hours and 20 minutes on a laptop, which is reasonable given the fact that there are 7 non-linear constraints. In Fig. 6.2 (orange dots), we show the focal lines for the seven design wavelengths. They show clear maxima at the desired focal length of 1.485 \(\mu\)m (marked by a yellow line). For a pure application perspective, we would use techniques such as successive refinement and multi-start to find a better optimum with less spurious foci. However, for the purpose of this paper, the current design shows a clear local optimum maximizing at the focal length for all wavelengths, which means that the optimization works with the surrogate model. The peak at the focal line at 9.99 \(\mu\)m wavelength clearly translates to a peak at a distance of \(\sim\)1750 \(\mu\)m at 8.57 \(\mu\)m wavelength and to a peak at 2000 \(\mu\)m at 7.49 \(\mu\)m wavelength.

6.2.2 Validation using local periodic approximation

In Fig. 6.2 (solid blue line), we show very good agreement between the approximate solver using the surrogate model and the LPA validation. It shows that the surrogate model can indeed be used for optimization purposes. And, optimizing for the surrogate model will also optimize for the online solver. The discrepancies are very difficult to catch with the naked eye, except for the wavelength 5.55 \(\mu\)m, where the solver-based surrogate model slightly overvalues the intensity at the focal point compared with the LPA validation. The intensity profiles along the focal axis all show—as desired—local maxima at the focal length in the LPA validation. This means that the optimization does not choose areas where the surrogate model would be inaccurate or non-physical (if there are any of those areas).
Figure 6.2: After optimization of a lens of 1 mm diameter with cylindrical symmetry (~85.5k unit cells with with 2475 degrees of freedom for 495 independent parameters) for [9.99, 8.57, 7.49, 6.66, 6, 5.45, 5.0] µm, the focal lines for the seven design wavelengths show clear maxima at the desired focal length of 1.485 µm (marked by a yellow line). The presence of multi-foci comes from the multiple constraints. The peak at the focal line at 9.99 µm wavelength clearly translates to a peak at a distance of ~1750 µm at 8.57 µm wavelength and to a peak at 2000 µm at 7.49 µm wavelength.

6.3 Conclusion

In this chapter, we proved that 1) a neural network can successfully increase the number of parameters per unit cell in the offline solver, and 2) optimizing using our inverse design framework based on this surrogate model optimizes solutions of Maxwell’s equations as well (as desired). Here, we showed an application of a surrogate model with six parameters. This would already be sufficient to implement a solver with overlapping domains in two dimensions with two parameters per unit cell. We think that neural networks have the capacity to extend the number of variables per unit cell to ≥ 10. In the next chapter, we present a method to actively learn a neural network in order to minimize the number of Maxwell’s equation solver queries necessary for the training phase.
Active learning of neural networks for electromagnetic surrogate models

In the present chapter, we present work from a collaboration with MIT-IBM Watson AI. We propose a neural-network surrogate model for Maxwell’s equations, predicting the scattering coefficients of a nanopatterned geometry as a function of many parameters that uses an active learning algorithm to greatly reduce the number of costly electromagnetic solves required in the training phase. Once our
surrogate model is trained, it can then be rapidly evaluated in order to design composite structures such as optical “metalenses” that combine millions of nanopatterned unit cells, turning supercomputer-scale electromagnetic design problems into laptop exercises. In contrast, we show that competing methods to construct surrogate models for electromagnetic problems (or other partial differential equations) typically require many more expensive training solves for large parameter sets because the training set is not adapted to the solution at hand. For active learning, we introduce a model for the surrogate uncertainty and we increase the training set in parameter regimes where the error is largest. We obtain good variance reduction and accuracy for subsequent optimization using the surrogate model, and we also show how the variance can be used in the surrogate model for exploration-exploitation “robust” optimization for improved performance in the face of uncertainty.

7.1 Motivation

In Ref. 3, we presented a framework of highly scalable inverse design in nanophotonics for metasurfaces—non-periodic patterned surfaces. The patterns are called unit cells and are generally (though not always) subwavelength; each unit cell has a discrete (1–3) number of parameters. This framework relies on a fast offline Maxwell’s equation solver, i.e., a surrogate model that—based on the parameters of the unit cell—evaluates the local field scattered by the unit cell under a local periodic approximation. Ref. 3 presents Chebyshev’s interpolation as an approximation technique to evaluate the local field function. The input of Chebyshev’s interpolation are Maxwell’s equations for the unit cells for a set of parameters on a Chebyshev grid. Therefore, the training set is $S_{\text{Chebyshev}} = \{x \in X, y\}$, where $|X|$ is the number of points needed to interpolate one dimension accurately and $n$ is the number of parameters in the unit cell. Although Chebyshev’s interpolation is proven to be the most accurate interpolation technique for smooth functions,\textsuperscript{14} this surrogate model scales exponentially with the number of parameters in the unit cell and the curse of the dimensionality applies when there are more
than 3 parameters per unit cell.

7.2 Main claims of this paper

In this paper, we present an active learning method to efficiently train a neural network for a surrogate model of Maxwell’s equations. The presented model both estimates the solutions and gives feedback on where it approximates poorly. We show evidence that our methods work successfully on a test problem where unit cells have 5 parameters. Our intuition why this type of approach works is that solutions of Maxwell’s equations are low-rank compared to the dimension of the input vector of unit-cell. The neural network efficiently capitalizes on the non-linearity of its activation functions to learn the underlying non-linear mapping which exploits this low-rankness, whereas Chebyshev’s interpolation does not.

7.3 Related work

Contrary to the purely exploration–exploitation problem and Bayesian optimization, it is fairly cheap to produce thousands of parameter–target pairs \(\{x, y\}\). Therefore, Gaussian processes’ regression\(^{115,116}\) will not scale well since the update is \(O(N^3)\), with \(N\) being the number of training pairs.

There is already a large literature about surrogate models, however to our knowledge, they fall into two categories: 1) models using unsupervised learning,\(^{117}\) and 2) models using regression paradigm (with homoskedastic error).\(^{118}\)

Supervised deep neural networks\(^{119,120,121}\) are usually used in a context where there is a lot of noisy training data. In our application, neural networks are used for their non-linear properties. The data are accurate because the solutions of Maxwell’s equation in the training set are exact. We aim at using the least evaluations possible for a given overall accuracy in the surrogate model.
7.4 Statistical model

7.4.1 Modeling uncertainties

Based on training data \( \{(x_i, y_i)\} \), where \( x_i \) are vectors and \( y_i \) are scalars, we model the target function and the uncertainty of the estimate with

\[
y = M(x) = \mu(x) + \sigma(x)\epsilon, \tag{7.1}
\]

where \( \mu(x) \) is the estimate of the function, \( \sigma(x) \) is the heteroskedastic variance, and \( \epsilon \sim \mathcal{N}(0, 1) \).

7.4.2 Neural network and objective function

We optimize the log-likelihood of our model\(^{122}\)

\[
-\log p(y_n|x_n) = \frac{\log \sigma^2(x_n)}{2} + \frac{(y - \mu(x))^2}{2\sigma^2(x)} + \text{constant}. \tag{7.2}
\]

In Ref.\(^{122}\), Lakshminarayanan et al. point out that this objective function might lead to overfitting. In the case of our surrogate model, this is a desired behavior because the evaluation of the functions is exact, so there are no outliers.

7.4.3 Ensembling

We train an ensemble of \( f \) neural networks to improve our estimates, \( M_i(x) = \mu_i(x) + \sigma_i(x)\epsilon \) for \( i = 1, ..., f \). We use the entire training dataset to train each network to improve the performance of the neural network. In order to de-correlate the prediction of the models, we set a different random seed for random initialization of the parameters, and implement random shuffling of the data points.\(^{122}\) As in \(^{122}\), we treat the ensemble as a uniformly-weighted mixture: \( M_*(x) = \frac{1}{f} \sum_{i=1}^{f} M_i(x) \) with mean
\[ \mu^*_i(x) = \frac{1}{J} \sum_{i=1}^{J} \mu_i(x) \] and variance \[ \sigma^*_i(x) = \frac{1}{J} \sum_{i=1}^{J} (\sigma^2_i(x) + \mu^2_i(x)) - \mu^2_i(x). \]

### 7.4.4 Active Learning

We use the ensemble of models to improve the training set so that the accuracy of our model is increased. We developed an active learning algorithm in which start with a random training set \( S_0 \subset \Omega \), where \( \Omega \) is the set of all possible input vectors of the unit cell. It is usually a hypercube because each parameter is bounded. We then successively add points to the training set to reduce the overall variance of the training set as shown in algorithm 2.

**Result:** \( \mu^*(x), \sigma^*(x) \)

n = 10000;

Sample \( X_0 \) uniformly in \( \Omega; |X_0| = n; \)

Compute \( \{y = Q(x), \forall x \in X_0\} \) to obtain \( S_0 = \{ (x, y), x \in X_0 \} \);

Train \( M_1, \ldots, M_J \) on \( S_0 \) and obtain \( M_\ast \);

for \( t = 1; T \) do

Sample 4n uniformly: \( X_{\text{ran} \text{dom}} \subset \Omega \setminus X_{t-1}; \)

Compute \( v_j = \sigma^*_i(x_j), \forall x_j \in X_{\text{ran} \text{dom}}; \)

Sort \( (v_i) \) and define \( X_t \subset X_{\text{ran} \text{dom}}; \) n points with biggest variance;

Compute \( \{y = Q(x), \forall x \in X_t\} \) to obtain \( S_t = \{ (x, y), x \in X_t \} \);

Train \( M_1, \ldots, M_J \) on \( S_{t-1} \cup S_t \) and obtain \( M_\ast \)

end

**Algorithm 2:** Active learning algorithm with an ensemble of neural networks to create a surrogate model of function \( Q \).
7.4.5 Results and discussion

Test unit cell. We consider a test unit cell with five layers in two dimensions. The first layer is a pole of silica with fixed height and variable width. The four other layers are air holes in the silica substrate with fixed height and variable width. The parameters are the width: \( w_i \in (0.1, 0.9) \), defined as a percentage of the unit cell period \( \Lambda \). We evaluate the function by doing a full Maxwell’s equation solve using Meep.\(^{123} \)

Qualitative and quantitative error reduction. We train two ensembles for the real and imaginary part of the scattered field, respectively; the architecture of the neural networks in the ensemble is showed in Fig. 7.1(A). This approach takes advantage of the smoothness of those functions compared to predicting the phase and amplitude of the field.\(^{109} \) We qualitatively assess the reduction of variance. We apply algorithm 2 with \( T = 1 \), and \( n = 10000 \), and \( \mathcal{J} = 5 \) on the ensemble corresponding with the real part to obtain the training set \( \mathcal{S}_{\text{train}} \). The ensemble for the imaginary part is then trained on \( \mathcal{S}_{\text{train}} \). In Fig. 7.1(B-D), we show the reduction of variance is more significant using our algorithm compared to choosing 10,000 additional random points.

In Fig. 7.2, we plot the fractional error of the neural networks’ estimates of the complex transmission on a separate test set. One neural network is trained using our active learning algorithm 2, and the other is trained by adding points randomly to the training set. The two network start with the same 10,000 data points. The active learning approach reduces the variance twice better, which shows early quantitative proof of concept. We also compare to Chebyshev’s interpolation—the state-of-the-art polynomial interpolation method for smooth functions. Active learning reduces the error by 7% by adding 10k points, which is more than twice better than adding 10k random points (2.5% reduction), and Chebyshev’s interpolation with \((7^5 - 6^5) \sim 9k \) additional points (2.8% reduction).
Figure 7.1: Fully connected neural network (NN) to approximate a low-rank function with three hidden layers containing fifty neurons each (A). (B) Estimates/Error vs targets with 10k training points. On the $x$-axis, we show the index of the points after sorting the test set w.r.t. to the target value of the real part of the local field. On the $y$-axis, we show the true real part (test set) in red, estimates from NN in blue dots, and error from NN (orange bar). The test MSE is 0.010. (C) Estimates/Error vs targets with 10k more random training points. The test MSE is 0.008. (D) Estimates/Error vs targets with 10k chosen using active learning. The test MSE is 0.004 which is twice better than the random set of additional points.

Validation of the surrogate model for large scale optimization. We want to use this surrogate model for large-scale optimization of metasurfaces. The surrogate model can be used to compute the intensity off a metasurface using the integral equation

$$|E(x)|^2 = \int_{\text{surface}} G(x, x') \hat{E}(x') \, dx'^2,$$  \hspace{1cm} (7.3)
Figure 7.2: Fractional error of the neural networks’ estimates of the complex transmission on a separate test set. One neural network is trained using our active learning algorithm 2, and the other is trained by adding points randomly to the training set. The two networks start with the same 10000 data points. The active learning approach reduces the variance twice better, which shows early quantitative proof of concept. We also compare to Chebyshev’s interpolation—the state-of-the-art polynomial interpolation method for smooth functions. Active learning reduces the error by 7% by adding 10k points, which is more than twice better than adding 10k random points (2.5% reduction), and Chebyshev’s interpolation with \((\gamma^3 - 6^3) \sim 9k\) additional points (2.8% reduction).

where \(E\) is the complex field and \(\hat{E}\) is the field scattered by the unit cells that the surrogate model computes. We want to optimize for the intensity at the focal spot. We are interested in the following questions: If we optimize using an objective function computed with our approximate solver based on this surrogate model, then does the optimum found also optimize the same objective function for any solutions of Maxwell’s equations? If the surrogate model has areas of inaccuracy which would make the solution non-physical (e.g., transmission with amplitude larger than 1), does our inverse design framework bring the optimal device to these areas? We use this surrogate model to optimize lens focusing a single wavelength consisting of 300 unit cells (1500 parameters) and with a numerical aperture of 0.32. We performed gradient-based optimization and obtained the gradient of the surrogate model using backpropagation in the ensemble of neural networks. In Fig. 7.3, we show the resulting intensity pro-
Figure 7.3: We performed gradient-based optimization. We obtained the gradient of the surrogate model using backpropagation in the ensemble of neural networks. We show the resulting intensity profile along the focal axis of the lens from the approximate solver based on the surrogate model (orange dots) and its local periodic approximation validation that uses an online solver for the unit cells (solid blue line). The approximate solver based on the surrogate model overestimates the intensity at the focal length, however it computes the profile correctly and has local maxima in the same places as the validation except for one spurious focus annotated by a red arrow. Since the maxima match well, an optimization using the surrogate model also optimizes solutions of Maxwell’s equation. Note that the discrepancy close to the metasurface (when \( z < 100 \, \mu m \)) comes from the local periodic approximation of the solver and not from the inaccuracy of the surrogate model.

file along the focal axis of the lens from the approximate solver based on the surrogate model (orange dots) and its local periodic approximation validation that uses an online solver for the unit cells (solid blue line). The approximate solver based on the surrogate model overestimates the intensity at the focal length, however, it computes the profile correctly and has local maxima in the same places as the validation except for one spurious focus, annotated with a red arrow. Since the maxima match well, an optimization using the surrogate model also optimizes solutions of Maxwell’s equation. Note that the discrepancy close to the metasurface (when \( z < 100 \, \mu m \)) comes from the local periodic approximation of the solver and not from the inaccuracy of the surrogate model. The fractional error of the
We performed gradient-based optimization and obtained the gradient of the mean surrogate model using backpropagation in the ensemble of neural networks for both the mean and the variance of our model. We show the resulting intensity profile along the focal axis of the lens from the approximate solver based on the surrogate model (orange dots) and its local periodic approximation validation that uses an online solver for the unit cells (solid blue line). Again, the approximate solver based on the mean surrogate model overestimates the intensity at the focal length, however it computes the profile correctly and has local maxima in the same places as the validation. Compared to the regular model, the robust model is qualitatively better, because it resolves for all the spurious foci, even the one annotated with a red arrow. Since the maxima match well, an optimization using the surrogate model also optimizes solutions of Maxwell’s equation. Note that the discrepancy close to the metasurface (when \( z < 100\mu m \)) comes from the local periodic approximation of the solver and not from the inaccuracy of the surrogate model.

"Robust" average optimization a.k.a. exploration–exploitation surrogate model

A common technique in optimization when the outcome is random is to optimize for a statistic of the outcome. Usually, optimizing for the worst case is more easily tractable. However, thanks to our

surrogate on the optimal parameters is higher than the fractional error on the test set, and is at 0.40.
neural network surrogate model, we can optimize for the mean

\[
\mathbb{E}|E(x)|^2 = \mathbb{E}\left(\int_{\text{surface}} G(x, x') \hat{E}(x') \, dx'\right)^2 \tag{7.4}
\]

\[
= \mathbb{E}(\int_{\text{surface}} G(x, x') \mu(x') + \sigma(x') \epsilon \, dx') (\int_{\text{surface}} G(x, x') \mu(x') + \sigma(x') \epsilon \, dx')^* \tag{7.5}
\]

\[
= \mathbb{E}(\epsilon^2) \left| \int_{\text{surface}} G(x, x') \mu(x') \, dx' \right|^2 + \mathbb{E}(\epsilon^2) \left| \int_{\text{surface}} G(x, x') \sigma(x') \, dx' \right|^2 + o \tag{7.6}
\]

\[
= \left| \int_{\text{surface}} G(x, x') \mu(x') \, dx' \right|^2 + \left| \int_{\text{surface}} G(x, x') \sigma(x') \, dx' \right|^2, \tag{7.7}
\]

where * denotes complex conjugates, and \( \epsilon \sim \mathcal{N}(0, 1) \), so \( \mathbb{E}(\epsilon) = 0 \), and \( \mathbb{E}(\epsilon^2) = 1 \). In the machine learning community, this type of approach is called an “exploration–exploitation” approach; the exploitation comes from the term depending on \( \mu \) and the exploration comes from the term depending on \( \sigma \). It encourages us to “explore” points where the estimator has a large variance.

We performed gradient-based optimization and obtained the gradient of the mean surrogate model using backpropagation in the ensemble of neural networks for both the mean and the variance of our model. In Fig. 7.4, we show the resulting intensity profile along the focal axis of the lens from the approximate solver based on the surrogate model (orange dots) and its local periodic approximation validation that uses an online solver for the unit cells (solid blue line). Again, the approximate solver based on the mean surrogate model overestimates the intensity at the focal length, however it computes the profile correctly and has local maxima in the same places as the validation. Compared to the regular model, the robust model is qualitatively better, because it resolve for all the spurious foci, even the one annotated with a red arrow. Since the maxima match well, an optimization using the surrogate model also optimizes for solutions of Maxwell’s equation. Note that the discrepancy close to the metasurface (when \( z < 100 \mu m \)) comes from the local periodic approximation of the solver and not from the inaccuracy of the surrogate model. The fractional error of the surrogate on the optimal parameters (with exploration–exploitation) is higher than the fractional error on the test set but lower.
than the fractional error of the previous design. The fractional error is at 0.36.

7.4.6 Conclusion and outlook

Although we currently have no proof of convergence, our algorithm seems to quickly capture where $M_\star$ needs more points to improve accuracy. We could validate the surrogate model in the inverse design framework described in Ref. 3. We currently use our error estimate only in the context of finding the points where our model performs poorly. However, we can also include it in our optimization framework, thus introducing an exploration–exploitation trade-off during the optimization.
In conclusion, the initial philosophical shift that my dissertation brought to the optics community is to frame the metasurface design problem as an optimization problem. In addition, my PhD work focused on design framework involving offline solvers. For the surrogate models—which act as offline solvers—I even set the problem as a data science problem. This thesis has surpassed the initial promise of its title, “Assume Your Neighbor is Your Equal” could be replaced by “Surrogate models,” which is not as attractive but goes beyond the local periodic approximation. Now that surrogate models can have up to tens of variables, we can use surrogate models for overlapping domains (multiplying the number of parameters per unit cells by 9 in three dimensions, and by 3 in two dimensions).

The problem of surrogate models was solved using a neural network, but we had a different paradigm than the neural network community. Andrew Ng, when describing machine learning, said the following: “An analogy is to building a space rocket: You need a huge rocket engine, and you need a lot of fuel. If you have a smaller rocket engine and a lot of fuel, your rocket’s probably not going to get off the ground. If you have a huge rocket engine but a tiny amount of fuel, you probably won’t make it to orbit.” Therefore, Andrew Ng implied that a neural network model needs a lot of data. Our approach to neural networks was diametrically opposed to this paradigm. We used a neural net-
work, but our goal was to train it with as little data as possible. Instead of having a lot of noisy data, we have an expensive solver that computes our training data exactly. Our active algorithm successfully takes advantage of neural networks’ versatility and non-linearity while also taking advantage of the exactness of the data we can query.

My dissertation should have a significant impact in the community because it started and contributed to bringing an optimization and data-driven approach to the design problem. The major contribution is to participate in bridging the gaps between the applied mathematics community, the applied physics community, and the machine learning community. This is beautifully exemplified in the collaborations I did during my PhD, which led to a optimization framework, tangible devices measured in a laboratory, and artificial intelligence algorithms. In the realm of optimization, my research has already spawned streams of research such as that of my colleague, Zin Lin. My original approach was disruptive because it enables one to design metasurfaces for any (discrete) function of the scattered field, whereas it used to be limited to only fully-known solutions of Maxwell’s equations. My research was also disruptive because it bridged the gap that there used to be between fabrication capability and design capability. Before, the fabrication capability was behind design capability. Now, it is matching the fabrication capability (centimeter lens from Chap. 5). Current extensions of my work in the topic topology optimization is even ahead of today’s fabrication capabilities, as are the multi-layer example unit cells from Chap. 7.

Going forward, I think it would be very beneficial for the research in fabrication methods and the research in design to start going hand in hand together, such as in the centimeter lens application of Chap. 5. Maybe new types of unit cells can be explored that can have more degrees of freedom while remaining easy to manufacture (for example: star shape unit cells, where the length of each branch would be a parameter). If we manage to add degrees of freedom (as shown in Chap. 6 and 7), the design methods presented in this dissertation will make metasurfaces ready for industry. Metasurfaces will meet the promise they made eight years ago: achieving better performance than traditional optics
while dramatically reducing fabrication costs. I think metasurfaces are ready to become an industry-
level technology.

For academic research regarding metasurfaces, the question of continuous objectives remains open
and will be the object of my future research. The present optimization framework still has to bridge
the gap between continuous objective (application specific, or industry standard) and the available
discrete objectives—function of the field.

Another forward direction for academic research regarding metasurfaces is to use the local periodic
approximation to optimize the coupling to a waveguide, which should possible as we found in Ref. 2
and partially solved in Appendix C.
To use standard high-dimensional optimization algorithms, one needs to provide an efficient computation of both the objective (cost) function and its gradient. There is a well-known technique called an adjoint method\cite{78} that can be used to efficiently compute the gradient for any number of parameters with a cost comparable to evaluating the objective function at most twice, which is commonly used in topology optimization.\cite{78,36,29,40,23,24,32,33} In the case of the two objectives presented in Sec. 3.3.1 and Sec. 3.3.2, the gradient is especially simple to evaluate as described in this Appendix.
In Sec. 3.3.1, \( f(p, s_0, \phi_0) = \int |s(p(x)) - s_0a(x)e^{i\phi_0}|^2 dx \), so the gradient is

\[
\begin{align*}
\frac{\partial f}{\partial p} &= 2\Re \left( \int (s(p(x)) - s_0a(x) \exp i\phi(x) + i\phi_0)^* s'(p(x)) dx \right) \\
\frac{\partial f}{\partial s_0} &= -2\Re \left( \int (s(p(x)) - s_0a(x) \exp i\phi(x) + i\phi_0)^* a(x) e^{i\phi(x)+i\phi_0} dx \right) \\
\frac{\partial f}{\partial \phi_0} &= -2\Re \left( \int (s(p(x)) - s_0a(x) \exp i\phi(x) + i\phi_0)^* is_0a(x) e^{i\phi(x)+i\phi_0} dx \right),
\end{align*}
\]

where \( \partial f/\partial p \) denotes the functional derivative\(^\text{124} \) with respect to the parameter function \( p(x) \) and \( * \) denotes complex conjugation. Notice that the computation of the gradient requires only the evaluation of a few simple integrals, comparable to the cost of evaluating \( f \). Similarly, in Sec. 3.3.2, \( g(p, x) = |E_z(x)|^2 = |\int_{y=y_0} G(x, (x', 0)) s(p(x')) dx'|^2 \), and therefore its gradient is

\[
\frac{\partial f}{\partial p} = 2\Re \left( \left( \int_{y=y_0} G(x, (x', 0)) s(p(x')) dx' \right)^* \int_{y=y_0} G(x, (x', 0)) s'(p(x')) dx' \right).
\]
Appendix to chapter 5
**Figure B.1:** Setup for measurement of the diode laser is collimated on the metalens. The scattered field is then collected by an objective lens, which enables us to measure the focus of a high NA lens. The light is recollimated and then focused on the camera. The measuring equipment is set on a translational piezoelectric stage to enable precise control of its position. The diode laser has a very narrow linewidth that is less than 1nm, while the linewidth of light from the SuperK is larger than 5nm. The bandwidth of light source will affect the depth of focus, as well as point spread function, so the diode laser provides better measurements. The drawback is that we have access to a more limited number of wavelengths.

**Figure B.2:** The intensity along the focal axis shows clear focusing at the focal spot for the six desired wavelengths which are linearly-spaced in the frequency domain to achieve best achromatic properties in the visible range: [490, 520, 540, 570, 610, and 650]nm. (A) The numerical aperture of this lens is 0.1; a diameter of 2mm corresponds to a focal length of about 1cm. The lens shows multiple spurious foci along the focal axis. Although the focal spot is clearly demarcated from the spurious foci, the spurious intensities can be bigger than the intensity at the desired focal spot. (B) The numerical aperture of this lens is 0.2; a diameter of 2mm corresponds to a focal length of about 5mm. This lens shows similar features as the NA=0.1 case. We plot the focal axis on a larger support compare to Fig. 5.9 to show the multi-foci.
Optimizing coupling to a waveguide

Building on work by Johnson et al., we want to use the local periodic approximation (LPA), to optimize the coupling between an incident planewave and a waveguide.

Theory of uniform case We start with this formula defining the coefficients from

\[
\frac{dc_m}{dz} = -\eta^*_m \sum_{n \neq m} \langle m^* | \frac{\partial}{\partial z} | n \rangle \frac{b_n(z) - b_m(z)}{b_n(z)} \exp \left( i \int^z [b_n(z') - b_m(z')]dz' \right) c_n,
\]  

(C.1)
where \( c_n \) are the coefficients of the expansion in waveguide modes. We neglected \(-\eta_m^s \langle m^s | \hat{B} \frac{\partial}{\partial z} \hat{e}_m \rangle\) as it only adds a phase term that we can set to zero.

The set of axes of reference used in the derivation is showed in Fig. C.1.

\[
\mathcal{E}(z)
\]

**Figure C.1**: Coordinate system for the derivation. We consider a waveguide in vacuum and want to optimize the permittivity function \( \varepsilon(z) \) to maximize the coupling of incident light to the waveguide mode. Following usual convention for waveguides, the \( z \) direction is that of the waveguide.

**Incident field and local uniform (periodic) approximation**  
We can separate the incident field \( |\psi\rangle \) into two components: one coming from the a locally-uniform field \( |\psi_0\rangle \) and the residual \( |\delta\psi\rangle \)

\[
|\psi\rangle = |\psi_0\rangle e^{ib_0z} + |\delta\psi\rangle
\]

\[
= \sum_{n \neq m} c_n^s |n\rangle + \sum \delta c_n |n\rangle \exp(i \int_{z'}^z b_0(z')dz'),
\]

where \( b_0 \) is constant and \( \exp(i \int z' dz' b_0) = \exp(ib_0z) \). The far-field incident wave does not contain guided modes so \( m \neq n \), where \( m \) are the guided modes.
Finding expansion coefficients Our goal is to calculate to first order in rate of change. We therefore will neglect the second term. Plugging decomposition C.2 into C.1, we get

$$\frac{d\epsilon_m}{dz} = -\eta^* \exp(-i \int b_m(z') dz') \left( \langle m^* | \frac{\partial \hat{A}}{\partial z} | n \rangle c_n \exp(ib_0z) \right)$$

$$\Rightarrow \epsilon_m = -\eta^* \int_0^L dz \exp \left[ i \int (b_0 - b_m(z')) dz' \right] \left( \langle m^* | \frac{\partial \hat{A}}{\partial z} | \psi_0 \rangle \right)$$

The decomposition is hard to find because it is a continuum. The term $$\langle m^* | \frac{\partial \hat{A}}{\partial z} | \psi_0 \rangle$$ simplifies (where we use the definition of $$\hat{A}$$ from 74). For two states $$\langle \psi_1 | = \begin{pmatrix} E_{1,t} \\ H_{1,t} \end{pmatrix}$$ and $$\langle \psi_2 | = \begin{pmatrix} E_{2,t} \\ H_{2,t} \end{pmatrix}$$, where p is changing in the z-direction, if we take $$\epsilon = 1$$ then

$$\langle \psi_1 | \frac{\partial \hat{A}}{\partial z} | \psi_2 \rangle = \langle \psi_1 | \begin{pmatrix} \frac{dp}{dz} \\ \frac{i}{w} \nabla \times \frac{dp}{p} \nabla \times \frac{dp}{dz} \end{pmatrix} \langle \psi_2 \rangle$$

$$= w \frac{dp}{dz} \int E_{1,t}^* \cdot E_{2,t} + \frac{1}{w} \int \nabla \times H_{1,t} \cdot \nabla \times \frac{dp}{p^2} \nabla \times H_{2,t}.$$  

From integration by parts, \( \int F \cdot \nabla \times G = \oint (\nabla \times F) \cdot G \), where the boundary term is zero since one of the two states is a guided mode. From Ampère’s law, $$\nabla \times H_{i,t} = +iwpE_{i,z}$$ for $$i=1,2$$. The tangential electric field gives the x and y-components, and the tangential magnetic field gives the z-component of the electric field, thanks to Faraday’s law and integration by parts. We get the simplified formula: $$\langle \psi_1 | \frac{\partial \hat{A}}{\partial z} | \psi_2 \rangle = w \frac{dp}{dz} \int E_{i,t}^* \cdot E_{i}$$ Therefore, the overlap integral is

$$\langle m^* | \frac{\partial \hat{A}}{\partial z} | \psi_0 \rangle = w \frac{dp}{dz} \int_{z,\text{cross-section}} E_{i,\text{MPB}}^* \cdot E_{i,\text{FDFD}},$$

since only the electric field in the y-direction is non-zero in the 2D case. Also for uniform and lossless waveguides, $$\hat{A}$$ is real symmetric so eigenvalues $$b$$ are real and $$E_i$$ is purely real74.
Therefore, the mode coefficient is

$$c_m = -\eta_m^* \int_0^L dz \frac{\int (b_0 - b_m(z')) dz'}{b_0 - b_m(z)} w \int_{z,\text{cross-section}} E_{y,\text{MPB}}^* \cdot E_{y,\text{FDFD}}. \quad (C.9)$$

We labelled $E_{y,\text{MPB}}^* \cdot E_{y,\text{FDFD}}$ with their corresponding numerical solver. MPB is a numerical solver to solve for the waveguide’s modes, solving Maxwell’s equations as an eigenvalue problem. FDFD stands for finite difference frequency domain; this simulation computes the effect of a source situated outside the waveguide in the frequency domain. The major difficulty of this method is to match those two solvers so that the cross-section integral is accurate enough to compute the coupling coefficient with local periodic approximation.
References


Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz


