METASURFACES FOR FIELD CONTROL AND DETECTION

by

Christopher A Rosenbury

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Optical Science and Engineering

Charlotte

2018

Approved by:

_____
Dr. Michael A Fiddy

_____
Dr. Glenn Boreman

_____
Dr. Tsing Hua Her

_____
Dr. Augustine Urbas

_____
Dr. Ryan Adams

ABSTRACT

CHRISTOPHER A ROSENBURY. Metasurfaces for field control and detection.
(Under the direction of DR. MICHAEL A FIDDY)

Metasurfaces have been of increasing interest in the scientific community due to their consistent ability to efficiently reproduce traditional optical elements via quasi-monolithic patterning. In this dissertation I will examine various metasurfaces and methodologies for designing them, explore a computational method for designing holographic metasurfaces via discrete iterative Fourier transform, and develop the proof of concept for a device capable of detecting electromagnetic radiation in a non-traditional manner. I demonstrate that full wave numerical solvers can act as a backbone for metasurface design, enabling a wide range of possible designs based upon dielectric resonator "building blocks" evaluated in computational solvers. Full wave simulations of metasurfaces designed from these building blocks are presented. I then utilize this backbone to form a multi-step inverse process which generates a design for a holographic metasurface based on any input image. This novel program enables fast and efficient holographic metasurface design through referencing a pre-filled database, populated by simulation of metasurface building blocks. I then extend the concept of simulation based design to a novel metasurface inspired electromagnetic field detector based upon the heterodyning of plasmonic waves generated by frequency selective meta-cells. This device will function as a metasurface plasmonic beat frequency rectifier, which will provide several advantages over traditional photodiode detectors. Some of these advantages include material design flexibility, direct wave detection which will give the possibility of phase information retrieval, and passive uncooled operation. A theoretical model, simulations, fabrication examples, and proof of concept testing will be included.

ACKNOWLEDGEMENTS

I would like to thank the following individuals: Mike Fiddy for his support and guidance throughout my work, Augustine Urbas for supporting my work and having me as an intern several times at the Air Force Research Laboratory which provided me with invaluable training and exposure to real world problems, Tsing Hua Her for many long conversations hashing out the physics at play in my work, Dean Brown for his help and support, Charles Rosenbury for providing assistance in database coding for the iterative Fourier transform work, and my wife Sarah Rosenbury for providing structural insight into the framing of this work and for the love support and encouragement while I wrote this. I would like to acknowledge the NSF I/UCRC Center for Metamaterials which provided funding for my work and introduced me to the needs and requirements of industry, as well as giving me the opportunity to network with various industry representatives.

DEDICATION

I would like to dedicate this dissertation to the almighty God and His son whom he sent to redeem us. It is through pursuing a greater understanding of His work that I developed many of the ideas presented in this document, and without whom I would not be half of the man I am today.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Metasurfaces have been hailed as the saving grace of metamaterials by virtue of their direct applicability, simplicity of implementation, and viability as commercial products compared to current bulk metamaterial research. Metamaterials are man-made structured materials with features smaller than the wavelength. With companies such as Kymeta and Google pursuing metasurface based solutions for difficult industrial problems, there exists a substantial interest and push toward developing these structures. As such there is an increasing need to understand, design, and invent new metasurfaces to meet commercial needs and interests. The field of metasurfaces

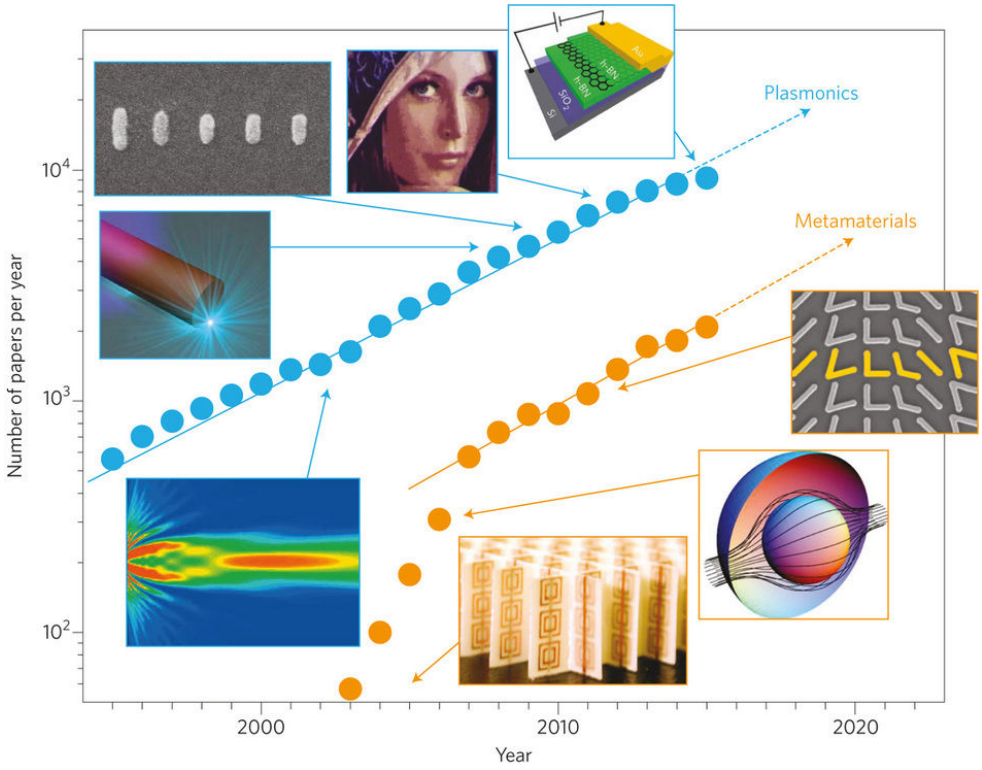


Figure 1.1: Publications per year in the fields of plasmonics and metamaterials. A substantial portion of the metamaterial publications in recent years are metasurface related. [1]

has been seeing an amazing growth in order to fulfill this growth in the market sector, experiencing substantial growth in publications and research (figure 1.1). I chose to direct omy focus on two areas that seemed particularly lacking in the the specific areas of design and invention for metasurfaces. First, for metasurface design I noticed that there was no standard method for the design of a holographic metasurface. Every research group approached the design of holographic metasurfaces with their own unique toolkit, and a common standardized approach seemed lacking. For the development of traditional holographic films, an industry standard is the use of iterative Fourier transform based designs. Capitalizing on this conceptual framework, I applied the concept to the development of a software which could quickly and cheaply design metasurfaces to produce an arbitrary field profile. This was accomplished by seeding a database with data from computational analysis of discrete meta-atoms to serve as unit cells for a comprehensive surface and utilizing this database to discretize the holographic metasurface design process. Second, for invention of new metasurface applications, I developed a metasurface inspired detector capable of detecting plasmonic beat frequencies via rectification, providing several advantages over current photodiode based detectors particularly in the mid and long wave infrared. These two products aim to improve and advance the field of metasurfaces through making these structures more accessible, easier to design, and compatible with current fabrication technology and thus more broadly appealing to industry.

1.1 Metasurfaces

Over the course of the development of the field of metasurfaces many definitions have been proposed to describe these structures. For the purpose of this work, I will use the following definition: metasurfaces are planar devices who's functionality is derived from utilizing sub-wavelength structure to control phase and amplitude. This definition, while broad, speaks to both the nature of the metasurface as a metamaterial and the functionality as a surface. These planar devices can be considered to

derive their functionality from three different regimes of physics. First, if the structure that gives the metasurface its functionality is sufficiently sub-wavelength (on the order of 10 times smaller than the wavelength in the material or less) then the metasurface will function according to effective medium theory. This results in a metasurface which operates in a manner similar to structures with a spatially varying refractive index, though more complicated effects can arise [2]. An example of this type of metasurface can be seen in figure 1.2 wherein the surface texturing imparts a geometric phase shift (also known as Pancharatnam-Berry phase) and thus a changing phase profile across the surface.

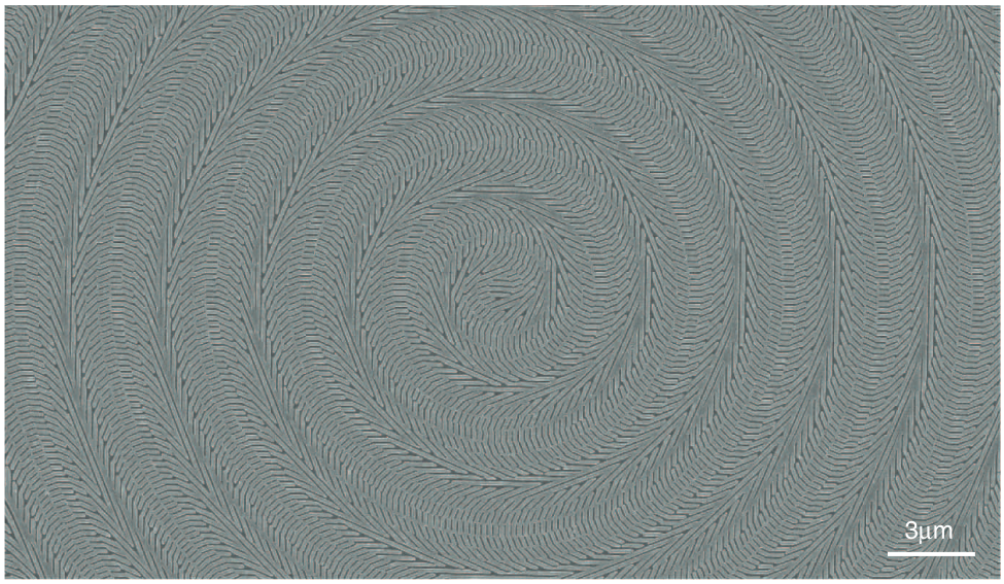


Figure 1.2: A SEM image of a fabricated non-resonant metasurface designed to function as an axicon. [3]

The second regime lies in metasurfaces that utilize discrete resonant elements. These elements typically have a size comparable to the wavelength in the material ($\sim\lambda/n$). These structures are used to discretely approximate a desired phase and amplitude value over the region of the resonators cross sectional absorption. By virtue of this, many optical elements can be substituted with metasurfaces operating in this manner, although optimization of these elements is an ongoing topic of research [4, 5, 6]. Designing metasurfaces which take advantage of this regime seems to be the

most common approach for designing metasurfaces. This may be due to the fact that utilizing discrete elements to approximate an analog distribution is fabricationally less challenging. This approach has its own challenges however. Because discrete elements are utilized, the field is not effectively continuous across the surface. These discontinuities tend to cause inefficiencies in the metasurface due to diffractive effects. While these effects can be minimized, it is difficult if not impossible to completely eliminate them. The third regime lies in the interaction of the resonant elements with each other, either as a nearest neighbor problem or due to a periodic arrangement of the elements giving rise to diffractive effects. This is typically considered as a secondary optimization point designed to improve device level efficiency.

The historical background for metasurfaces started as early as 1902 with the discovery of the Woods effect leading to plasmonics [7]. From there, metasurface development was inevitable. Some of the earliest structures that could be considered as metasurfaces were designed during WW2 to aid in radar lensing [8]. Development of structured materials continued with the field of frequency selective surfaces, which as a broader category included all periodic structures comparable to the wavelength in order to selectively affect one particular frequency over another. While the earliest examples of these structures dated back to 1919, work on them peaked in the 1980s [9]. These structures tended to be designed for microwave frequencies, with applications ranging from radomes to filters. Applications and research ranged over many material types, although dielectrics were prefered for trasmission applications and metals were preferred for reflection. The structures were then generalized to higher frequency applications and became known as photonic crystals. Work continued on the crystals pushing the lower size limits until the structures were indistinguishable from the modern definition of metamaterials. However, it was not until recently that these structures became known as metasurfaces and gained a coherent research effort. This was largely due to the work of Frederico Cappasso and his seminal paper “Light

propagation with phase discontinuities: generalized laws of reflection and refraction" in which he established a firm methodology for creating beam deflection via a surface phase profile generated by discrete resonant elements [4] (figure 1.3). Since then, the field of metasurfaces has continued to expand, with the majority of work focusing on resonant elements discretely representing analog phase and amplitude distributions.

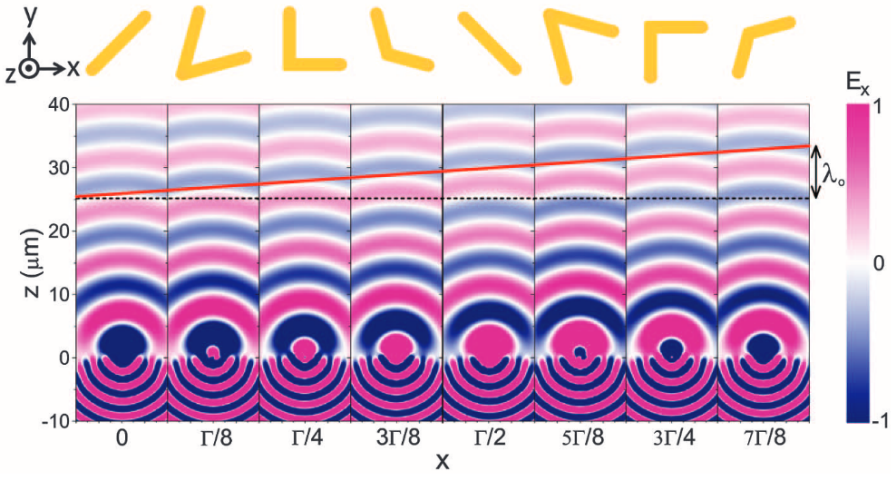


Figure 1.3: A string of plasmonic elements across a surface generating a phase gradient resulting in beam deflection. Here Λ represents the full unit cell dimension. This phenomena combined with the discretized execution resulted in the beginning of the field of metasurfaces. [4]

Metasurfaces have many different applications, both for resonant and for nonresonant type surfaces. These applications encompass essentially anything that can be accomplished via standard optics, as metasurfaces are simply a discretized method of controlling amplitude and phase in a "small" thickness. These structures can typically be divided into three different categories based on the type and shape of the resonant elements. The simplest case is the "2D" case, where the resonators can be manufactured by a single step lithography process. In other words, the resonators have vertical side walls. Examples of such structures are cubes and vertically oriented cylinders. The next case is called "2.5D" to indicate it is halfway between a two dimensional and a three dimensional case. These structures could be approximated by a multistep lithography process, but cannot be fabricated in a single step except by specific

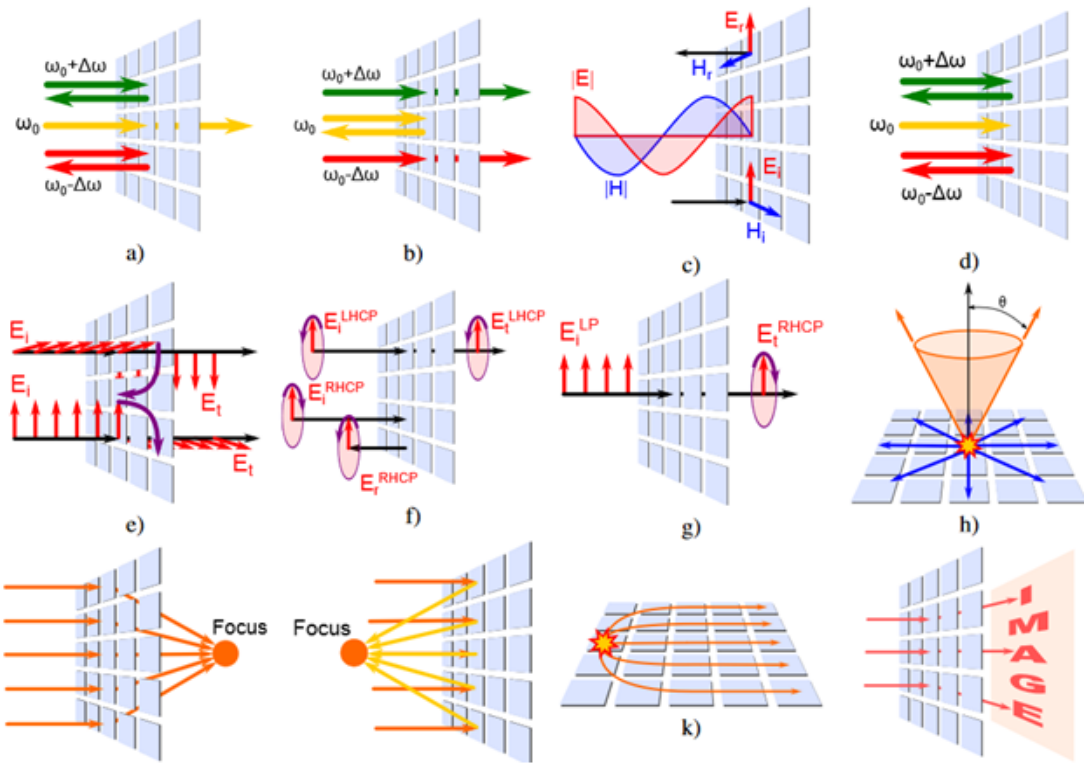


Figure 1.4: a) Bandpass frequency selective surface, b) Bandstop frequency selective surface, c) High-impedance surface, d) Narrowband perfect absorber, e) Twist polarizer, f) Right-handed circular-polarization frequency selective surface, g) Linear-to-circular polarization converter, h) Two-dimensional leaky-wave antenna with a conical-beam pattern, i) Focusing transmit array, j) Focusing reflectarray, k) Flat Luneburg lens, l) Hologram [6]

instruments. These shapes no longer have vertical side walls, but have monotonically decreasing cross sectional area with increasing height. Examples of these resonators include hemispheres, cones, and pyramids. The final case of "3D" resonators include all resonators with varying cross sectional area with increasing height. These shapes cannot be fabricated using traditional lithography, which makes the practical implementation of such surfaces difficult. Examples of such resonators range from something as simple as a full sphere to a small scale replica of the statue of liberty. The dimension of the resonators tend to be on the order of $\lambda/2n$ for refractive index n of the material. This allows for a large number of elements to be placed in a relatively small area, providing an exellent approach to discretize traditional optics allowing a

smaller form factor and lighter weight solution. Some examples of the applications which can be replicated by metasurfaces are shown in figure 1.4. In this figure, each of the squares represents a metasurface unit cell operating upon the incident radiation to affect a specific outcome. While this list is not comprehensive, it does indicate that metasurfaces can be used for a substantial number of applications which are currently being filled by traditional optics. Indeed, since the form factor is so small, many of these functions can be scaled down to operate on a much smaller size profile than traditional optics would otherwise allow.

1.2 Plasmonic Metasurfaces

A specific subset of metasurfaces which is useful to address for this work in particular is the topic of plasmonic metasurfaces. These are metasurfaces for which the primary functionality is due to Surface Plasmon Polaritons (SPPs). In order to properly address what SPPs are, let us first examine the nature of plasmons and polaritons. A plasmon is an oscillation of electrons in a plasma with well defined frequency and wavelength. As such, these waves are longintudinal and are traditionally excited with neutron beams or electron beams due to the requirement of a large quantity of momentium required to excite them [10]. Polaritons on the other hand are coupled dipole oscillations with electromagnetic radiation. Polaritons can have many forms, such as phonon polaritons, exciton polaritons, and of course SPPs. Due to the confinement to the surface region, the required momentum for generating a plasmon polariton at the surface is reduced. This allows electromagnetic waves couple with electron oscillations on the surface of a free electron gas. The matching of momentum, called the dispersion of a surface plasmon is given by

$$k_{sp} = k_o \sqrt{\frac{\epsilon_m \epsilon_d}{\epsilon_m + \epsilon_d}} \tag{1.1}$$

for waves along the direction of propagation of the SPP, where k_o is the free space wavevector given by ω/c, ϵ_m is the permittivity of a metallic layer, and ϵ_d is the permittivity of a dielectric layer. Since this condition is very sensitive to the material properties, the condition for generation of a SPP at a surface can be narrow in the frequency domain. This phenomena is known as the surface plasmon resonance and is used to determine slight changes in the dielectric properties of a surface for sensing purposes [11]. When this effect is reduced to a small spatial feature such as in nanoparticles, it is referred to as a localized surface plasmon resonance. SPP's will be dealt with in more detail in Chapter 3 Section 3.1, but for now it is enough to understand that plasmonic metasurfaces operate via the utilization of materials with a free electron gas at the desired frequency (such as metals for optical frequency), a structure designed to couple free space waves into SPPs by satisfying equation 1.1, and a geometry designed to utilize these waves for the localized control of the electromagnetic field.

For many plasmonic metasurfaces, the inclusion of plasmonic elements are simply another method of locally altering the phase and amplitude of the field, such as from figure 1.3. Each individual element acts similar to an antenna, rebroadcasting the field with a specific phase delay. This approach to plasmonic metasurfaces is best used for reflecting surfaces, as the transmission efficiency is typically low due to the large metallic content on the surface. However, this is not the only approach for utilizing plasmonics in metasurfaces. One application, and one that is specifically useful with regards to the development of the detector, is the area of perfectly absorbing metasurfaces. These surfaces utilize stacked plasmonic waveguides to act as resonant cavities for a broad band of frequencies (figure 1.5). The stack is tapered so that each waveguide is resonant to a slightly different frequency. This results in broadband high efficiency absorption. Essentially, the surface waves resonate inside the layers as waves would resonate inside a Fabry Perot cavity, losing energy to the losses of the

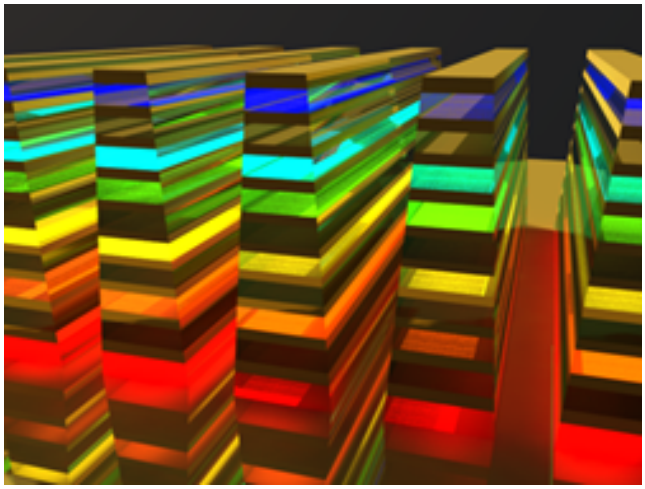


Figure 1.5: A visual example for the frequency selective resonances in a perfectly absorbing metasurface. [12]

metallic layers. Because there are so many elements in such close proximity and at high coupling efficiency, a significant portion of the field is absorbed.

CHAPTER 2: METASURFACE DESIGN

There are many and varied methodologies for designing metasurfaces, and each approach depends highly on the application. Some things, however, are held in common among design methods. To begin with, all resonator based metasurfaces must have a dimension specific amplitude and phase response of the resonators to be used in the design. While these responses can be obtained many different ways, the two predominant methodologies are via direct theoretical computation (such as Mie theory for dielectric scatterers [13]) or simulation via a numerical modeling software such as COMSOL, Lumerical, or HFSS. The second approach seems to be more prevalent as access to this software and computational power is becoming easier to obtain.

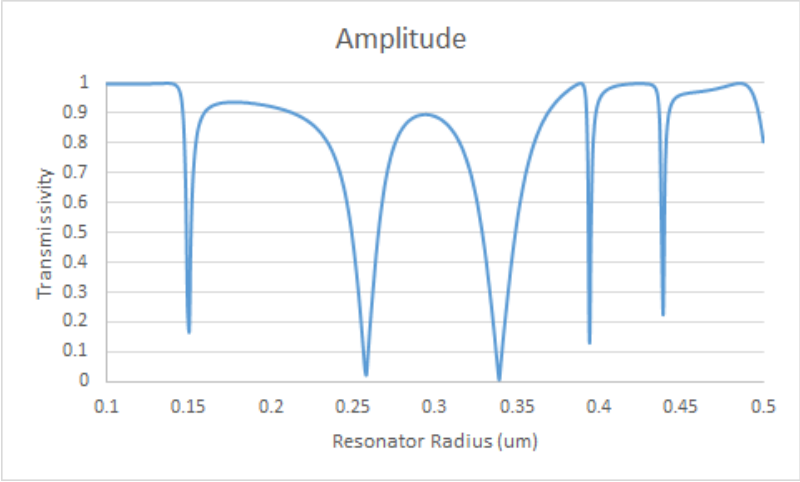
The first consideration in designing a metasurface is the phase curve of the resonator, as without the appropriate topological phase the metasurface will have minimum functionality. This is due to the fact that interference is governed predominantly by the phase of the field, and as such the phase profile will be the major contributor to the far field pattern on propagation. An example amplitude and phase curve is shown in figure 2.1. Ideally, this curve will encompass a full phase interval, such as $0 - 2\pi$ (or $\pi - 3\pi$, etc.), with a gradual transition that gives lighter tolerances for fabrication. In practice, it can be difficult to find a single resonator that will satisfy this as many resonances tend to be quite sharp or only span π.

The second thing that metasurface design has in common is the choice of phase profile. While for certain applications of metasurfaces this process can be complicated, many of the simple optical elements have a well known near- field phase profile that gives rise to the response. A parabolic mirror, for example, establishes a phase $\phi \propto r^2$

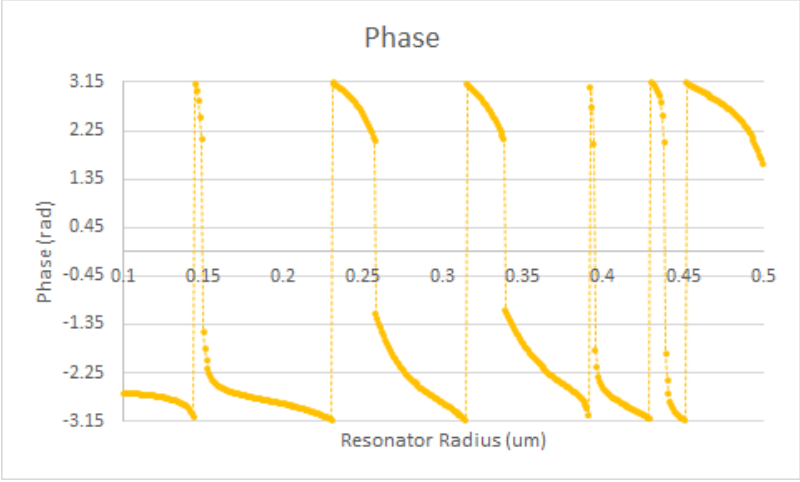
for some distance r away from the center of the mirror. In addition, the modified Snells law

$$n_1 \sin\theta_1 + n_2 \sin\theta_2 = \frac{1}{k_o}\frac{d\phi}{dx} \tag{2.1}$$

tells us that by establishing a phase gradient, it is possible to obtain beam deflection. These are but a few examples of what can be accomplished with the appropriate design of a phase profile. While these simulations were done for narrow band illumination,



(a) Amplitude Response



(b) Phase Response

Figure 2.1: An example amplitude and phase curve for a resonant structure. This was obtained from a cylindrical resonator over varying radius. By choosing the correct radii it is possible to establish a phase profile across a surface through assembling these structures.

simultaneously satisfying the phase profile conditions for multiple frequencies will provide a response that is effectively broadband [14]. This type of advanced profiling and others of its kind that rely on arranging large datasets typically function off of an optimization algorithm. While there are many ways to develop specific phase profiles for metasurface application, I will look at two in this work. First, utilizing a numerical simulation to obtain a phase profile for a resonator and constructing metasurfaces via these resonators. Second, leveraging a Fourier transform approach in conjunction with a database of resonators (established via methodology from the first approach) to design metasurface holograms.

2.1 Design Via COMSOL: Focusing Elements and Deflection

One method for establishing a metasurface design is through direct computational methodology. This approach works well for matching to an analytical phase distribution, as the arrangement is set by a known spatial phase profile. The first step in developing a metasurface from a computational methodology is to determine the software that will be used in the numerical calculations. While there are many pieces of software capable of giving us the information required (HFSS, C-Star, etc), I have chosen to use COMSOL multiphysics due to the inherent flexibility in design space that COMSOL offers. The second step is to obtain a unit cell from which an amplitude and phase profile can be assembled. The main feature of the unit cell is the resonator, for which a shape must be chosen. I chose to pursue cylindrical resonators, which were appropriate to the application the designs were performed for [13]. Since the intended operation wavelength was 0.97 microns, I chose to use GaAs as the material due to the high refractive index of 3.5227 which reduces the overall size of the elements, allowing for a smoother discretized phase profile. In order to obtain effective amplitude and phase response for the resonator, I needed to establish a methodology for extracting the parameters from the simulation. In order to do this, I will use the well accepted practice of Scattering (S) parameter retrieval. For this

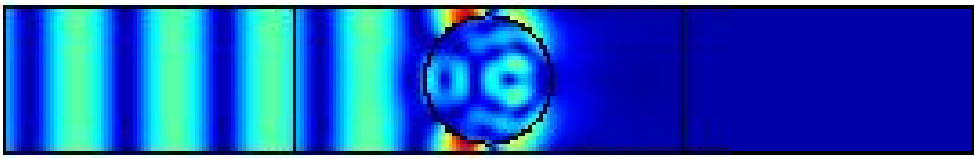


Figure 2.2: COMSOL simulation of GaAs cylinder resonator. Ports are on both sides, and periodic boundary conditions on top and bottom. Field excitation is from the port on the left.

particular simulation I assume periodic boundary conditions to simulate an array of resonators, which allows us the extra parameter of nearest neighbor spacing to adjust thereby broadening the design space. I use port excitation and the full field solver to obtain meaningful S parameter data. S parameters in this case are defined by

$$S_{21} = \frac{E_2}{E_{1_o}} \tag{2.2}$$

which is to say the ratio of the field at port 2 due to excitation from port 1. Thus S_{11} would be the ratio of the field at port 1 due to excitation at port 1. Although these values are technically averaged over the entire port, the size of the ports relative to the resonator will give a good indication of the response normal to the surface. Now, it is fairly straightforward to realize from this that the amplitude of this value will be the transmitted field or the reflected field strength. The phase however, is not so obvious. If the input is a plane wave of the form Ae^{ikx}, I am able to extract the phase response of the resonator, since

$$S_{21} = \frac{A_2}{A_o} e^{i\phi} \tag{2.3}$$

which gives us the relative phase ϕ. For the purposes of metasurface construction, if you are using the same resonator for all of the elements, the relative phase is sufficient. If you are utilizing multiple different resonators though, then the absolute phase response needs to be determined. In order to extract the actual phase from the calculated S parameter, I must assume that the scattered field detected by the

ports approximates a plane wave (which is often a reasonable assumption in the small angle limit). Thus I can divide the fields into component phase parts given by their k vector multiplied by the traveled distance. For example, $E_{1_o} = A_o e^{ikx_1}$ and $E_2 = A_2 e^{i(kx_2+\theta)}$ for the setup in figure 2.2. When the S parameter is then solved for theta, the propagation terms will essentially act as a baseline shift for the phase. Extracting the absolute phase is not needed for the designs in this section, but is used in the next section to establish a database of resonator responses. An example of this simulation is seen in figure 2.2.

Now that I have established a methodology by which to obtain amplitude and phase response from a resonator, I need to fill out the design space by conducting a sweep of the radius and the resonator spacing (typically better done as a sweep of the period and fill factor so as to keep the results in a full space rather than a half space) to obtain a range of amplitude and phase responses. Though it may not be immediately obvious, the resonator spacing plays a significant role in the net response due to the coupling and multiple scattering when multiple resonators are in

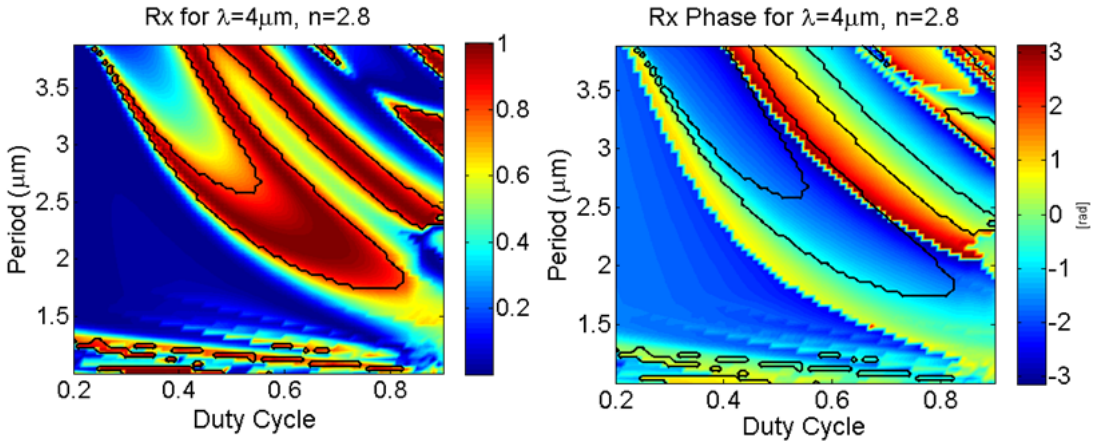


Figure 2.3: The full reflectivity and phase spectrum as a function of period and duty cycle (fill percentage) for the cylindrical resonator as shown in figure 2.2. The selected regions are areas where the reflectivity lies above 80%. By comparing these regions on the phase diagram, it is clear to see that a full 2π phase shift can be achieved whilst maintaining higher than 80% efficiency.

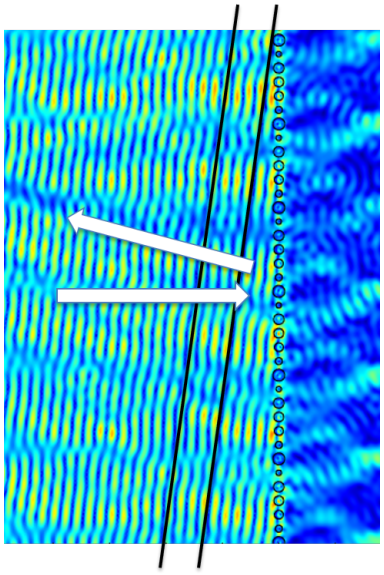
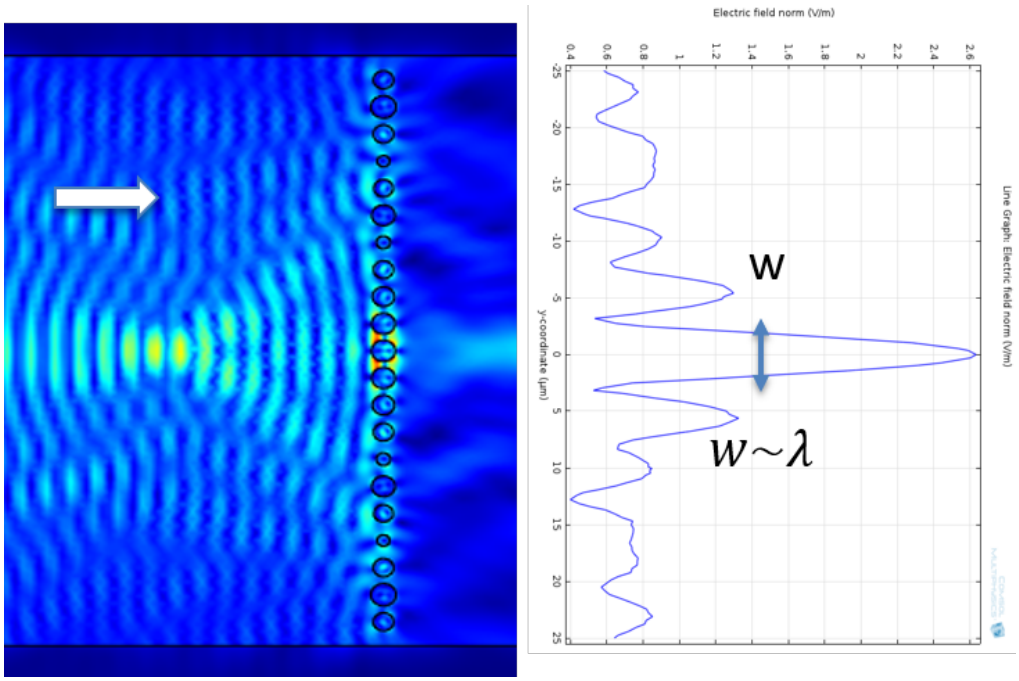
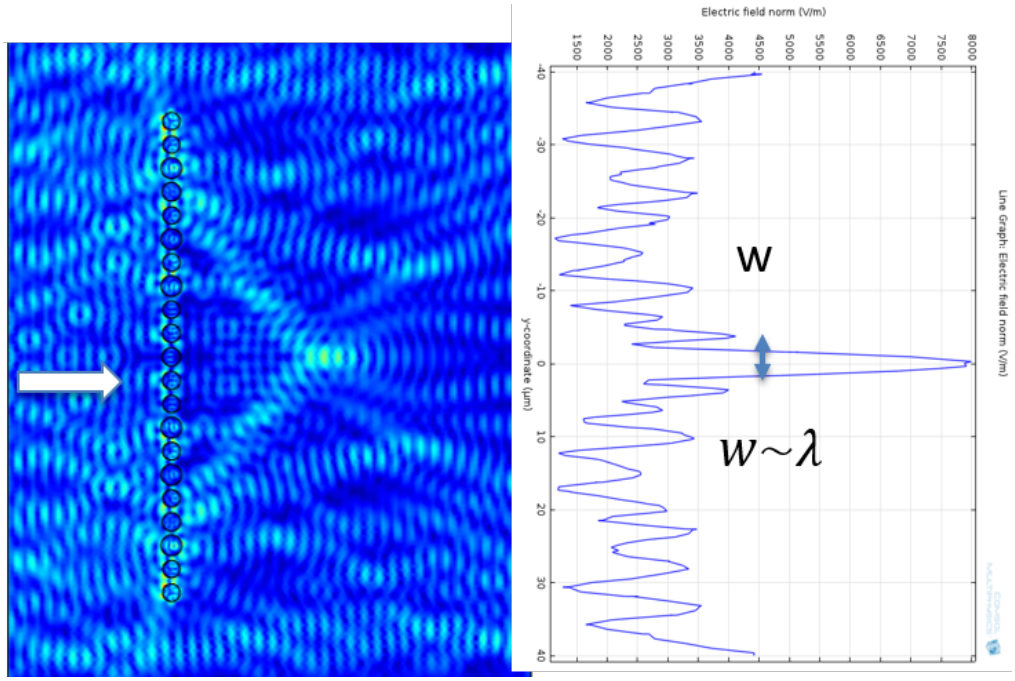


Figure 2.4: A beam deflecting reflector constructed of cylindrical resonators, with resonators placed to satisfy the modified Snells law for a 15 degree deflection.

close proximity. This provides an extra degree of freedom that allows us to construct a full design space for the the resonator response. Figure 2.3 shows the full reflectivity response of the simulated resonator over the field of availability. This gives us a full design space to work with. With this design space determined, I can now pick and choose resonators to fit a spatial phase profile that I desire. For example, if I want to get beam deflection I can set up a phase profile according to equation 2.1. I then repeat that unit cell multiple times to create a large array that will cause beam deflection. A simulation of this can be seen in figure 2.4. Although the reflected field in this case is not quite uniform, further optimization of the exact placement of the resonators can improve this. I can also match the phase profiles for a lens or a focusing mirror (figure 2.5). These flat lenses and flat focusing mirrors can have a very high optical power or a low optical power depending on the design. The lens and focusing mirror displayed are quite high power, but also have a good resolution with the spot size on the order of the wavelength.

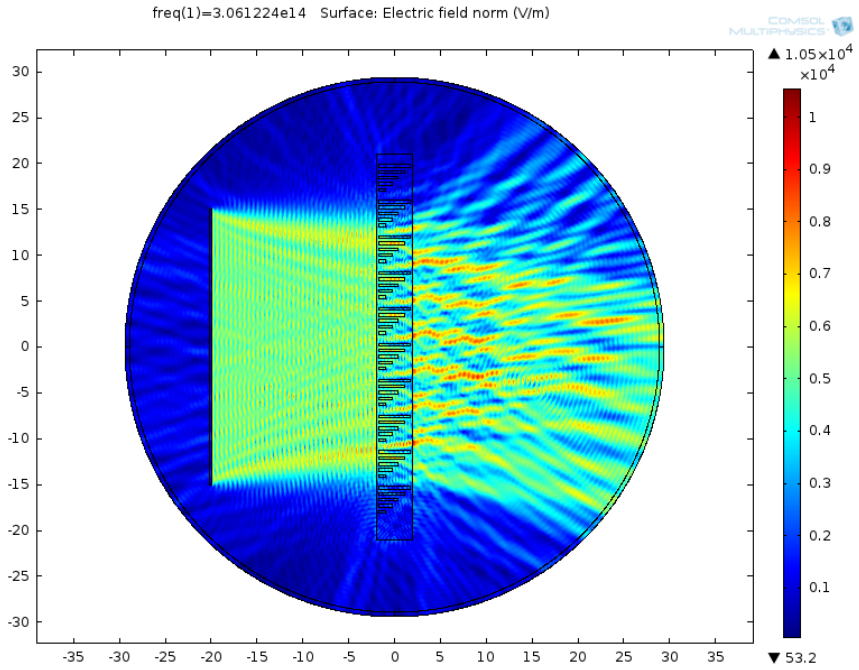


(a) Focusing Mirror and Spot profile

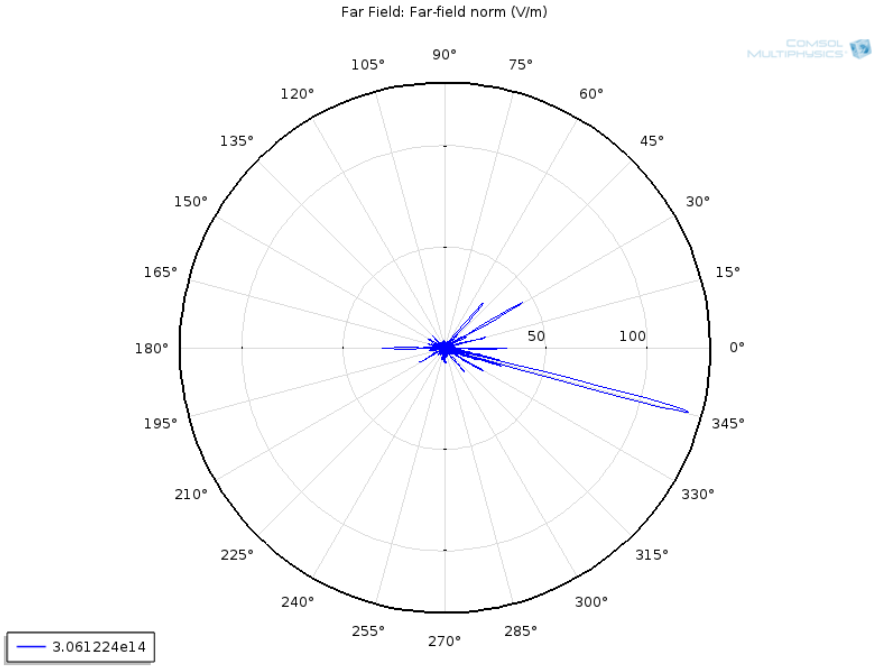


(b) Metalens and spot profile

Figure 2.5: (a) Focusing mirror metasurface with reflectivity of 90% and the corresponding field profile through the focal spot. (b) Lensing metasurface with transmissivity of 80% and the corresponding field profile through the focal spot.



(a) Deflecting sheet full simulation



(b) Far field pattern of deflecting sheet

Figure 2.6: (a) Beam deflecting metasurface from voids in material (b) Far field scattering pattern of the deflecting surface. Although there is some scatter into other modes, the predominant energy flow is at 15 degrees. This deflector is approximately 3 wavelengths thick

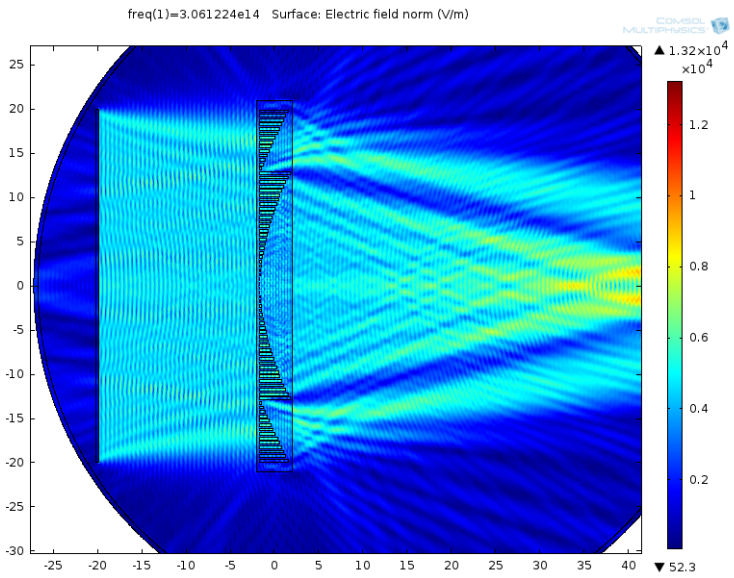


Figure 2.7: A lens made from holes in material. The focal spot was designed at 60 microns, which is far to the right of the image. This is similar in look and design to the inverse of a Fresnel lens, but could be embedded in a glass to provide protection.

I also considered different approaches that didn't strictly rely on strictly resonant effects in order to improve the bandwidth of operation. One thing I considered was to utilize structured holes in a material to provide the neccesary phase profile. Since the requirement for a metasurface is the relative phase difference between adjacent elements, a lack of material can function in a similar manner to the presence of material. The phase shift is obtained from a strategically sized void in a material, whereupon the lower index of the void will provide a phase shift similar to (but opposite) the presence of a resonator in air. Utilizing this methodology of strategically placed voids in glass, I was able to construct simulations that provide deflection and lensing in transmission (because the electromagnetic wave must pass through the void to gain the phase shift). These results are shown in figures 2.6 and 2.7. These simulations indicate that voids in materials can act in a similar manner to surface texturing. With the ability to generate bubbles in glass from localized laser focusing, fabrication of metasurface features from occlusions and defects becomes a very real and manufacturable process.

I have demonstrated through rigorous simulation that metasurfaces can function as beam deflectors and microlenses with significant focusing power. I have shown that these surfaces perform with reasonable efficiency and have acceptable focal properties while reducing size and weight. These structures are but a sample of the available options for designing metasurfaces via simulation, and many more applications have been explored by other research groups.

2.2 Design Via Discrete Fourier Transform

Simulation via rigorous simulation is well and good, but designing a metasurface via this methodology is time consuming and computationally expensive. While, as noted earlier, there are methods which take advantage of algorithms for improving design, these approaches tend to be either limited in scope or computationally intensive. Genetic algorithms iterating on a specific resonant element to modify a specific field profile can be quick, but can only produce variations upon what it was designed for (such as a lens in reference [15]). These methods utilize discretization, a well known resonator response, and an analytic phase profile to optimize a structure that will produce the desired response. While this is an excellent approach to reducing the computational expense and time required for design, creating a new or modified algorithm for every desired field pattern would likely be costly in terms of man hours and development without an advanced artificial intelligence driving the effort. As such, a generalized approach for designing metasurfaces for arbitrary input would be ideal. In following with a similar design methodology to phase hologram design, utilizing a Fourier transform based approach opens the door towards this goal.

The basic property of Fourier transforms that I will utilize for this work is the occurrence of Fourier transform upon propagation. A Fourier transform will occur naturally at a distance d beyond the surface so long as the condition,

$$\frac{L^2}{d\lambda} \ll 1 \tag{2.4}$$

is met, where L is the largest dimension of the diffracting element. This is known as the Fraunhauffer diffraction far field [16]. This is particularly useful because it demonstrates that I can utilize Fourier effects without the need of lenses and special systems (like 4f correlators). Thus, if I begin with a field profile that I wish to create, I know that the inverse Fourier transform will produce the near field response that would create the desired pattern upon propagation. For the purposes of this work, I will consider the near field to be the region within $\lambda/2\pi$ of the surface of the radiating element, consistent with the reactive region of electromagnetically short antennae.

In order to properly take advantage of the nature of this process, I developed a computer program written in Java and referencing to a mySQL database of resonator values. The idea was that an arbitrary digital image could be uploaded, an inverse Fast Fourier Transform of the image could be done, and the resultant matrix of near field data values could be mapped to resonators from the database, giving a full engineering diagram for a metasurface that would produce the input image. In order to accomplish this task, certain hurdles had to be overcome. First, a fast Fourier transform only works on images which have a pixel dimension that is a power of two. Thus, any input image would need to be cropped and/or padded to be able to execute a transform. This is most easily accomplished by starting from the top left corner of the input digital image matrix, comparing the dimension to choose the nearest power of two, and truncating and filling as needed.

Second, if I begin with the complete knowledge of the magnitude and phase that I wish to create an image of, this process can easily generate the distribution of magnitude and phase needed for a metasurface to create the desired image. There is however, an inherent difficulty associated with this approach. Typically the phase profile of a desired image is not typically known beforehand, at least in higher frequencies. As most detectors simply measure magnitude, the phase information is typically lost. In order to properly create a Fourier transform based approach, first

a method to recover a phase profile consistent with the image must be developed. This is mostly due to the fact that the majority (∼90% [17]) of the information of an image is contained in the phase profile and thus the stability and quality of the image generation is highly dependent on the phase. A constant phase will introduce noise into the transformed array, requiring a very large dynamic range difference between pixels, increasing diffractive effects. A random phase on the other hand reduces the dynamic range and thus diffractive effects, but introduces speckle into the image. The ideal balance between these is to pick a phase that is appropriate for the individual image. Thus, if I want to create efficient holograms, I must encode both the amplitude and the phase in such a manner that the resultant pattern produces the desired image. Whilst phase reconstruction typically requires an iterative algorithm, I can take advantage of the fact that I am discretizing the representation to limit the available options for possible phase choices.

In order to simplify the process of phase retrieval I first realize that after the inverse transform, the mapping to the resonators must map at a specific precision level. There are two main justifications for this level of precision: first the physical fabrication limitations will prevent a perfect match to the transform, and second there is an impracticality to creating a database populated with enough resonators to match every possible response perfectly for a large variance in materials and wavelengths. In addition, although some machines may be able to fabricate with higher precision maintaining a decimal limit would provide better tolerances for the machines in question, ensuring repeatability and quality over multiple manufacturing runs. For the purposes of this work I have set the limit to two decimals of precision, as manufacturing limitations will likely preclude a finer precision (at least at higher frequencies where the phase argument is actually of concern). For example, a machine designed to fabricate an average dimension of 1 micron would need to be accurate to within 10nm, which is far more achievable than being accurate to within 1nm. With this

in mind, I return to the original image cognizant of the fact that I will likely not do better than two decimals of precision in recreation of the image. Thus, I can employ a phase retrieval by method of best fit. For any magnitude, there are an infinite number of combinations of real and imaginary values which produce that magnitude, and at an infinite number of angles. However, since I can only fit to two decimals of precision, not every combination will produce the desired input magnitude. To choose, first take all possible two digit real values, square them and subtract them from the magnitude squared. After taking the square root, this gives us the list of the imaginary values which correspond to the real.

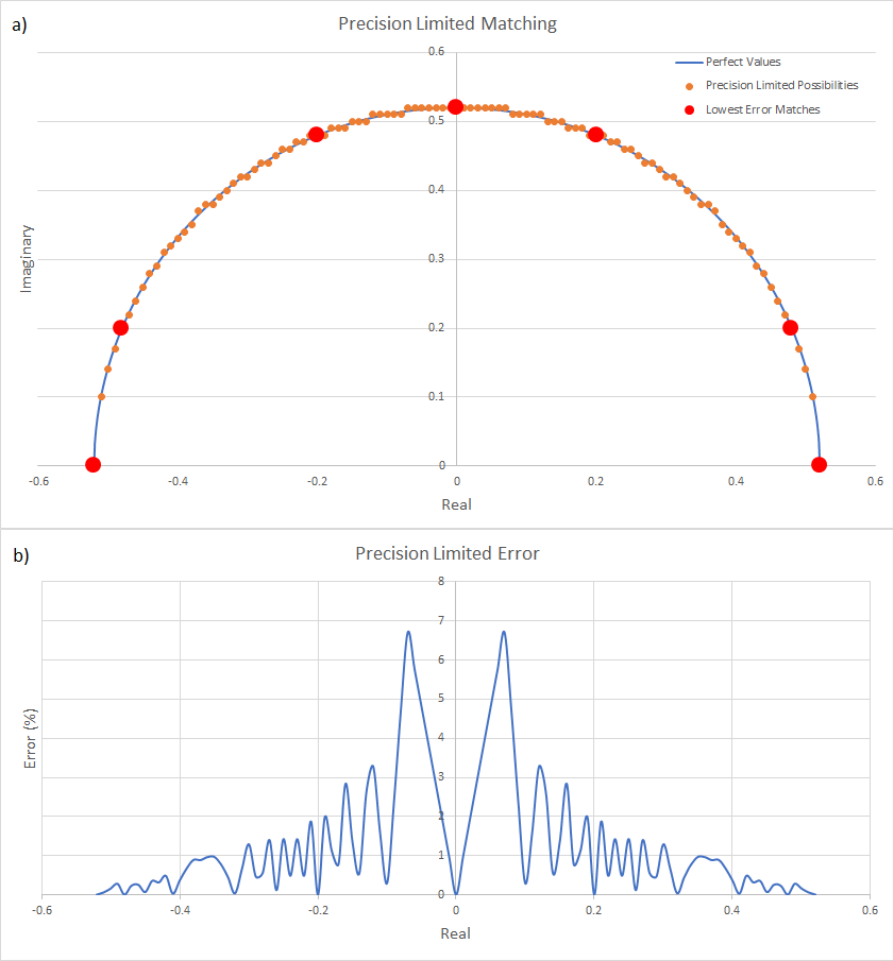


Figure 2.8: a) Upper plane of possible values for a given magnitude, the precision limited choices, and the best fit values. b) Percent error of the precision limited generated magnitudes compared to the actual values.

I then truncate the list of imaginary values to two decimals, and find the resultant $magnitude = \sqrt{real^2 + imag^2}$ and compare to the original magnitude value. The best fit is the chosen phase (figure 2.8). However, there is still some degree of ambiguity to this, because I do not know which of the matching points or which of the four quadrants the phase is in due to the methodology of obtaining these numbers. At this point in time I utilize a random number generator to select which quadrant the phase lies in, but this can be refined into an iterative phase retrieval method such as the Gerchberg-Saxton algorithm [18] (which is still being improved for better holographic generation [19]) later on to retrieve an even better phase. Once a phase is established, I then take the inverse Fourier transform which results in a series of values that can then be mapped to the metasurface resonators. A database is constructed of resonators which contains the extracted S-parameters of the resonator response, the physical dimensions (length and width for rectangular resonators, radius for cylindrical resonators, nearest neighbor spacing, and height with respect to the surface), the wavelength the parameters were extracted for, whether the result was reflective or transmissive, and the intended angle of polarization. Orientation of width is defined along the 0 degree polarization angle. For the time being the intent is to design polarization independent metasurfaces, so focus for populating the database has been on polarization independent shapes such as squares and cylinders. For two decimal precision matching, there are at most ten thousand resonators needed for the database per critical parameter.

HEIGHT	IMAGINAR	IS REFLECT	LENGTH	LOADERID	LOADERKEY	MAGNITU	MATI	Unit	SPACING	PHASE	Pol	RADIUS	REAL RE	WAVELEN	WIDTH
0.75	-0.21	1	0.5	483	3573c436-7	0.21095	Si	microns	0.72	-1.4758	0	-1	0.02	1.5	0.5
-1	-0.15	1	-1	19	3573c436-7	0.15811	Si	microns	-1	-1.8925	-1	-1	-0.05	1.5	-1
-1	-0.11	1	-1	143	3573c436-7	0.1118	Si	microns	-1	-1.7506	-1	-1	-0.02	1.5	-1

Figure 2.9: A selection of three database entries for the metasurface design database. The first entry represents a complete resonator, whereas the second and third are incomplete. When designing resonators to match, you can match either the real and imaginary from S parameter values, or the magnitude and phase. The program currently only supports rectangular resonators (length and width), although cylindrical resonators can easily be included (radius).

However, each image is likely only to use a small subset of those resonators. In a database populated by only three thousand resonators (likely with some duplication), an example fitting required just six hundred resonators to match to the inverse transform for a 1024 square pixel image. The parameters required for input into the resonators can be seen in figure 2.9, where the database is in a .csv format. In order to ensure that coupling is not a problem, each metasurface pixel is designed to contain at least 25 resonators, ensuring the resultant near field amplitude and phase are accurate representations of the database value.

2.3 Results

(a) Original input image

(b) Resultant reproduced hologram intensity

Figure 2.10: (a) Original image input into the metasurface design program (b) Left: Resultant image magnitude predicted from the generated metasurface at one decimal of precision. The pattern can barely be made out. Right: The resultant image magnitude predicted from the generated metasurface at two decimals of precision. A much cleaner image

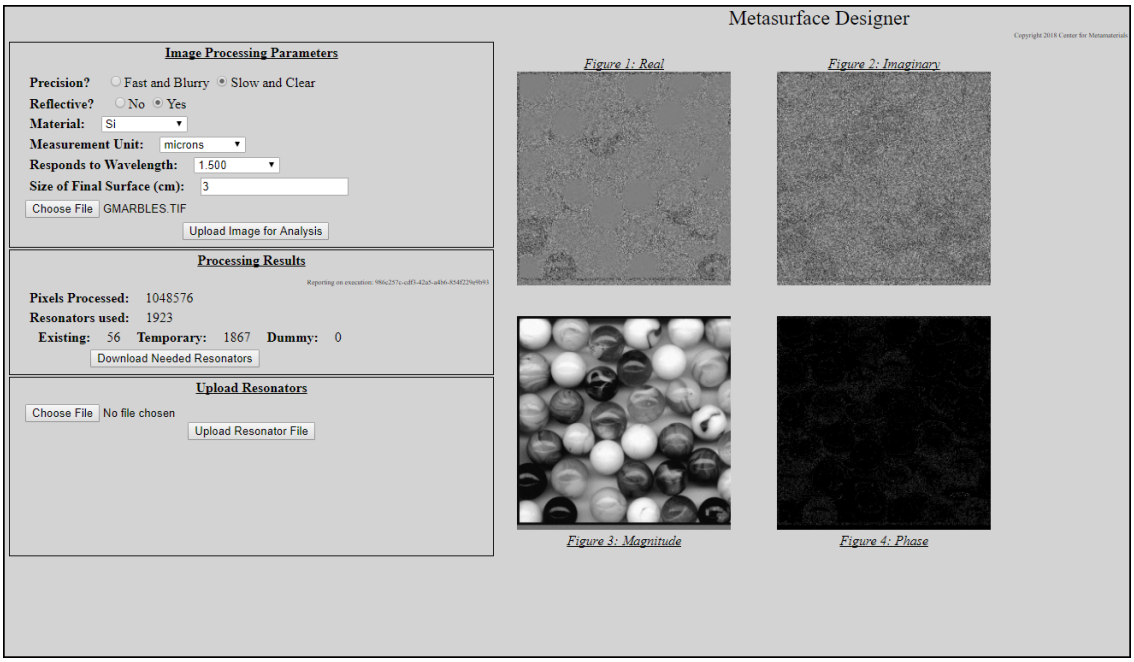


Figure 2.11: The GUI for the metasurface design program with an image having been processed.

Once the program was functional and could assign a phase to the input image, take the inverse transform, assign resonators, and transform back forward to produce the expected result (as seen in figure 2.10) there were only two things left to do. The first was to design a graphic user interface (GUI) which would properly accept an input image, execute the transform process, and output either a list of resonators that needed to be included in the database or output an engineering diagram for use in physically generating the metasurface. The GUI required inputs to determine the type of metasurface designed. Selectable options needed to include whether the desired metasurface would be reflective or transmittive, what the desired operating wavelength was, the desired material, and the overall desired size of the metasurface. Other included options were a choice of either one decimal precision (fast an blurry) or two (slow and clear) and the measurement units (to make transitioning to RF metasurfaces easier). The GUI also displays the real and imaginary values for the resultant transform as well as the magnitude and phase profiles. Upon completion

of the transform, if there were resonators that were not contained in the database, the option to download a .csv file with the required parameters appears. Once downloaded these parameters can be determined, filled in to the spreadsheet, and uploaded to update the database. The program can then be run again and will provide the option to download the engineering diagram. The GUI can be seen in figure 2.11. I then

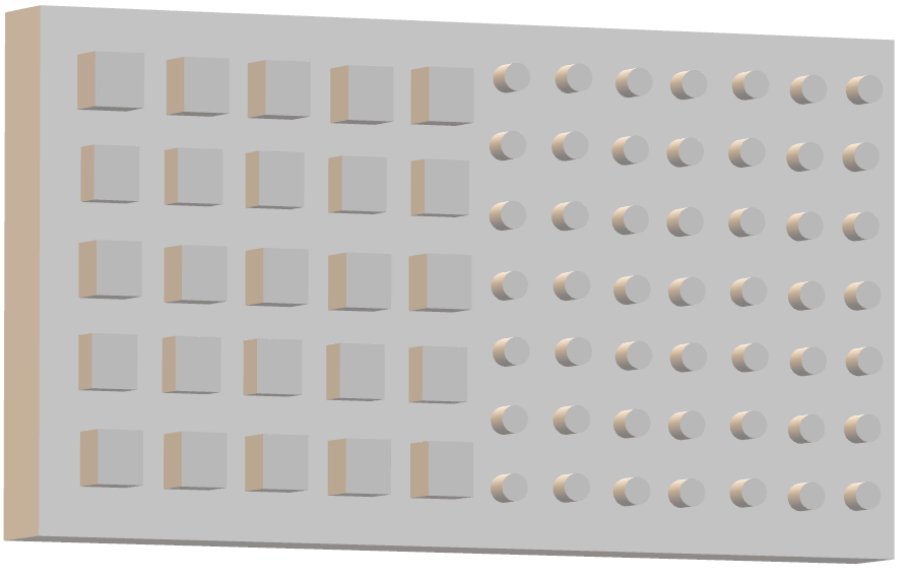


Figure 2.12: A example of two metasurface pixels.

needed an engineering diagram format which allowed us to represent the resonators which had been mapped to pixels to a schematic which could be used in manufacturing. While there are many different formats to choose from I decided to follow the .dxf format as it is fairly easy to code and is an open source file format. Because I do not utilize scaling and rely on a one to one mapping of pixel size to image size, this results in metasurfaces that can be rather large, with .dxf files in the gigabyte file sizes. In order to reduce this, further improvements on the program may include switching to a .step file, which is the ISO standard. This file type is more difficult to use, but when implemented properly can significantly reduce the file size without loss of information. Figure 2.12 shows an example of what two pixels might look like from this approach. This program provides the first step to a database driven holographic

metasurface design through Fourier transform algorithms.

CHAPTER 3: DETECTOR THEORY

The wide diversity in applications of metasurfaces opens up many new applications that were either not possible before or were too difficult to accomplish in previous years. Deep inspection into the field gives rise to a greater understanding of electromagnetic interactions at surfaces, which can lead to new perspectives and applications of existing principles. It was through the study of the perfectly absorbing metasurface and noticing the similarities in structure between the metal dielectric stacks and the outer segment of cone cells (figure 3.1) in the human eye that I first developed the idea of a metasurface inspired detector. Although the similarities are predominantly superficial, it was enough for us to envision a methodology by which the metal dielectric layers may serve as a detector. This approach, separate from the concept of designing metasurfaces to replace optical elements, is focused around opening up an entirely new metasurface related field through direct field detection.

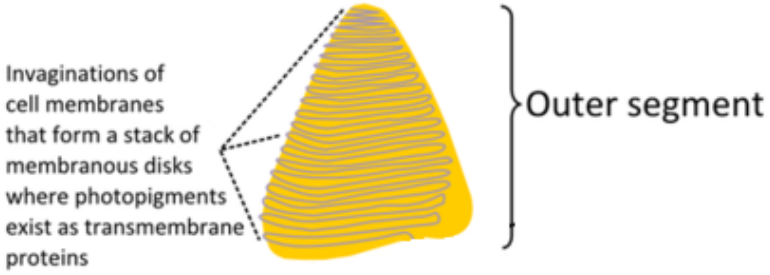


Figure 3.1: A representative diagram of the outer segment. Chromophores acting as transmembrane proteins exist between the folds of membranes. These chromaphores, once excited, will travel down to the stem and register an absorption event.

The initial idea was to harness the plasmons that formed in between the layers as the active region of the detector. However, since the plasmons themselves were too high frequency to be directly detected by traditional electronics, I envisioned uti-

lizing the multiple resonances (figure 3.2) to create a beat frequency of interference between two layers. The first concept was to read out the beat frequency from the layers by channeling the plasmons down a contact on the side of each pillar (figure 3.3). Attempts to realize this approach are fraught with challenges, predominantly due to the interference of the contact layer with the establishment of the resonances in the layers. I was unable to envision a solution for readout in this configuration which addresses the practical difficulties of reading out the signal without interfering with the fundamental operation without substantially altering the design. Thus in order to develop a design which is compatible with a traditional manufacturing process and would function as a detector operating on the principle of plasmonic beat frequency interference, I must look at the physics in a more detailed way and establish a methodical design beginning from basic principles.

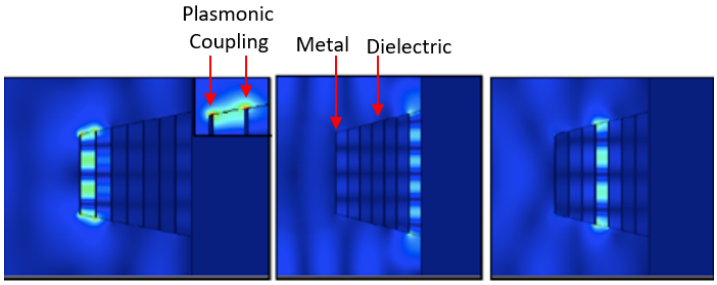


Figure 3.2: COMSOL simulation of the stacked plasmonic waveguide absorbing structure. The coupling due to scattering at the metal edges and resonances at three different frequencies can be clearly seen. Although there is some bleed-over, the field is mostly confined to a single layer.

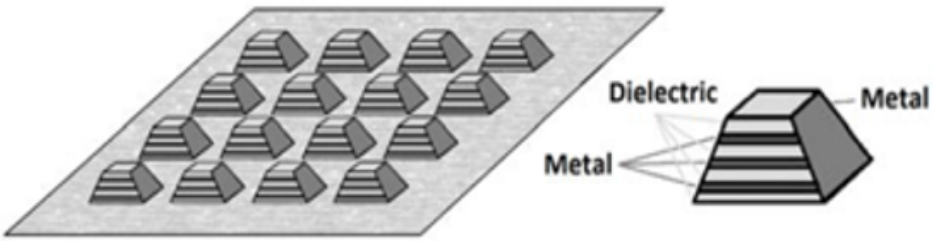


Figure 3.3: A visualization of a focal plane array of detectors utilizing the initial idea for readout.

3.1 Plasmons and Beats

To begin with, let us consider the basic physics and mathematics of SPPs and beat frequencies. Although I could begin from a rigorous solution of Maxwell's equations to establish the existence of SPPs it is rather tedious and has been done many times before, to the point where it was well established even in 1988 [20]. Sufficient to say, the electric field of a SPP is given by the equation of a standard monochromatic electromagnetic plane wave (shown here in two dimensions propagating in the $\hat{\boldsymbol{x}}$ direction)

$$\boldsymbol{E} = A_o e^{i(k_x x + k_y y - \omega t)} \hat{\boldsymbol{x}} \tag{3.1}$$

with the sole difference of the wave vectors k_x and k_y. For a simple surface with the normal in the $\hat{\boldsymbol{y}}$ direction, k_x is given by Equation 1.1, wheras k_y is a tad more complicated since it depends on whether you are interested in the field in the dielectric or field in the metal. The k vector for the field in the dielectric is given by

$$k_{y_d} = k_o \epsilon_d \sqrt{\frac{1}{\epsilon_m + \epsilon_d}} \tag{3.2}$$

whereas the vector in the metal is given by

$$k_{y_m} = k_o \epsilon_m \sqrt{\frac{1}{\epsilon_m + \epsilon_d}} \tag{3.3}$$

This relationship is particularly important when it is realized that for a metal with a magnitude permittivity greater than the magnitude permittivity of the dielectric, these equation are purely imaginary (a relationship assumed . Thus, when substituted back in to equation 3.1, it results in an exponential decay away from the surface. Example field vectors can be seen in figure 3.4.

With SPPs properly defined, now let us look at beat frequencies an how a beat in a SPP might appear. Consider two waves, one at frequency ω_1 and the other at

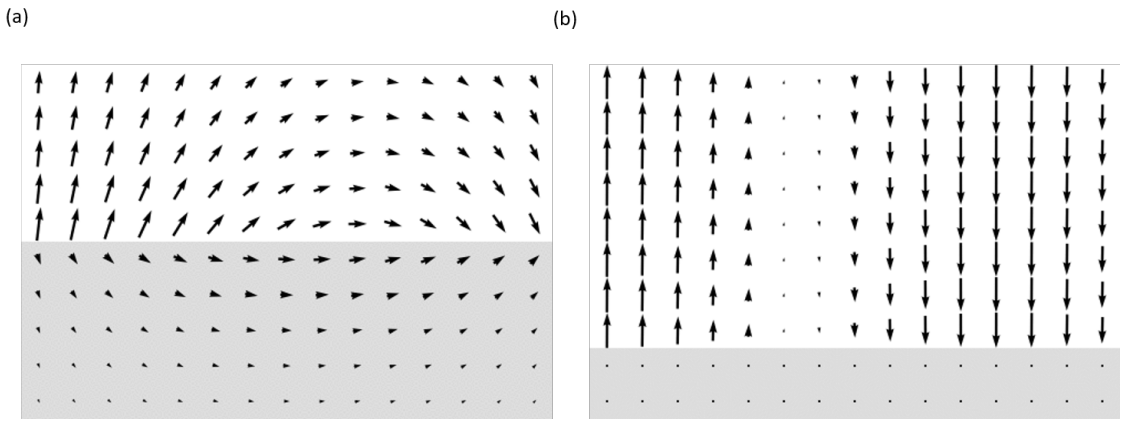


Figure 3.4: Electric field vectors for plasmonic waves at an air metal interface with (a) low magnitude metal permittivity (-2.6 + 0.6i) and at (b) high magnitude metal permittivity (-2700 + 1600i). Of note, the field penetrates much further into the metal for a low magnitude permittivity.

ω_2. In any linear medium, the field in the region of interference will be given by

$$\boldsymbol{E} = A_1 e^{i(k_{1x}x + k_{1y}y - \omega_1 t)}\hat{\boldsymbol{x}} + A_2 e^{i(k_{2x}x + k_{2y}y - \omega_2 t)}\hat{\boldsymbol{x}} \tag{3.4}$$

which by grouping terms and a little bit of clever mathematical manipulation I can turn that into

$$\boldsymbol{E} = A_1 e^{i(k_{1x}x + k_{1y}y - \omega_1 t)}\left(1 + \frac{A_2}{A_1} e^{i[(k_{2x}-k_{1x})x + (k_{2y}-k_{1y})y - (\omega_2 - \omega 1)t]}\right)\hat{\boldsymbol{x}} \tag{3.5}$$

for which it is relatively clear that there is a core central oscillation at ω_1 and an envelope in both space and time given by the difference between ω_1 and ω_2 arising from the relationship between k and ω given by equation 1.1. A spatially varying beat (which is directly analagous to a temporal beat by the above equation) with the envelope can be seen in figure 3.5. Of a particularly interesting note is that this equation is predominantly temperature independent. The permittivity (and thus each k vector) will change very slightly with temperature which may, depending on the dispersion of the thermal shift, result in the beat frequency changing very slightly. Given that our intention is to have a very small difference between frequencies

however, it is highly unlikely that the effect would be measurable except for extreme or exotic materials and thus I will consider the plasmonic interference to be thermally independent. This treatment only deals with coherent radiation, but the requirements to obtain interference in the partially coherent case will be introduced later. Since my goal is to rectify the temporal beat, the question arises, what is the maximum power that can be rectified if I can only rectify the beat? The answer is fairly straightforward. Since power is the time integral of the field squared, I know the power in the beat is directly proportional to the modulation depth squared. Specifically

$$P_b = P_1 * \frac{m^2}{4} = P_1 \frac{\left(\frac{A_2}{A_1}\right)^2}{4} \tag{3.6}$$

where P_1 is the power in signal 1 and the factor of four is due to the fact that we are looking at a single sideband. Thus, at perfectly equal amplitude the maximum rectifiable power from the beat frequency is 25% of the total power. I can therefore say

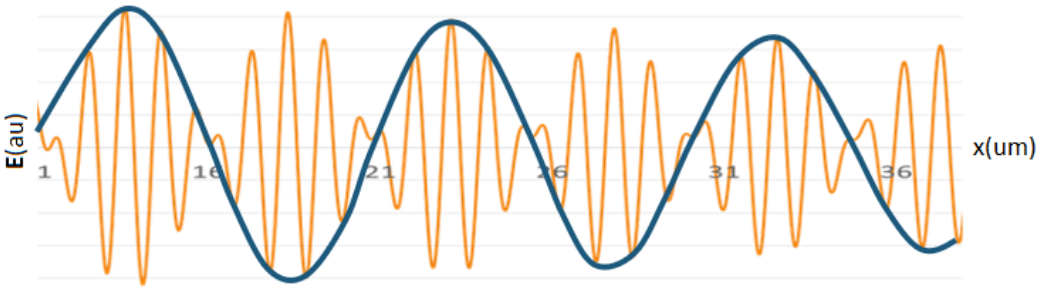


Figure 3.5: Example field with actual field values oscillating in space in orange, and the effective envelope in blue.

with great confidence that plasmons, if they are coherent and interfere, will generate a rectifiable beat in both space (interference fringes) and time (beat frequency). This is actually a fairly well established phenomena in the literature, with research being conducted in this area for many years [21, 22, 23]. Many applications for utilization of spatial beats has studied, including sensors [24], quantum information [25], nanolithography [26], plasmonic electromagnetic induced transparency [27], and many on

chip application, some of which have resulted in patents ([28],[29],[30]). In addition, many of the applications of plasmonic interference can be directly tied to work in metamaterials and metasurfaces [1]. As early as 1996 cursory studies were conducted with plasmonic beat frequencies using a laser and a spatial light modulator to provide a near adjacent frequency and interfere the plasmons to detect a beat [21]. This picked up again in 2011 when a group used optical heterodyning on a grating to produce a measurable beat frequency of about 1GHz [22]. Most recently in 2016, a group using a Zeeman effect HeNe laser was able to utilize polarization modulation to amplitude modulate the interference signal from plasmons generated on a surface [23]. These studies in particular indicate the physical feasibility of detecting plasmonic beat frequencies, and open the door for designing a detector based on these principles. There are however, many things to take into consideration when designing a device to function as a detector utilizing these basic principles.

3.2 Device Considerations

The goal of this section is to utilize plasmonic beat frequencies to create an integrated on chip detector for electromagnetic radiation. I wish to achieve this by rectifying the beat frequency for a readout by traditional electronics. By utilizing plasmonics instead of electron-hole pair generation, I gain several key advantages including an increased customizability due to material availability, sensitivity to the relative phase of the contributing frequencies, and a passive operation that can be biased for improved signal to noise ratio. In addition, the active region will not suffer from photobleaching or over exposure (though the rectifying diode may still suffer from those difficulties). In order for us to design a complete device, five main elements are needed: the intended light source, the choice of material, the method of launching plasmons, the rectifier, and the readout. The most critical choice for these elements is in the choice of light source. This is predominantly due to the fact that in order for interference to occur (either spatial or temporal) the detector/diode must

be within the coherence length of the plasmon launching source. To get a better idea of how coherence affects the fundamental device interaction, let us take a closer look at coherence.

Coherence is a term used to compare how similar signals are to each other. This can refer to either two different signals or a single signal compared to itself at a different point in time or space. There are two basic types of coherence: temporal coherence and spatial coherence. Temporal coherence typically relates how similar a signal is to a time delayed version of itself. This is directly related to the bandwidth of the signal, as multiple frequencies will interfere to reduce the similarity of the signal to itself after a period of time. This period of time is known as the coherence time and is inversely proportional to the bandwidth.

$$t_c \approx \frac{1}{\Delta\nu} \tag{3.7}$$

A graphical example of this can be seen in figure 3.6. Spatial coherence has two different components, the first being coherence length. Coherence length is similar to coherence time in that it is the distance that the radiation can travel within the coherence time. The second component is coherence area, which is typically what people refer to when referencing spatial coherence. Coherence area is the area over which one wavefront is virtually identical to the next. While this varies by source, it can be substantially improved by passing the radiation through a pinhole, which by Hyugens principle will create spherical wavefronts. These spherical wave fronts will have a very high degree of spatial coherence. Thus, if I were to start with a very incoherent source such as a light bulb, I could improve the spatial coherence by introducing a pinhole or some other form of spatial filter; and I could improve temporal coherence by introducing a chromatic filter to reduce the bandwidth (figure 3.7). This will downselect to a narrow frequency and spatial region, which will only

be a fraction of the total power emitted by the incoherent source, but this could be partially mitigated by engineering the surface which is detecting the radiation and focusing the radiation onto the sensitive regions.

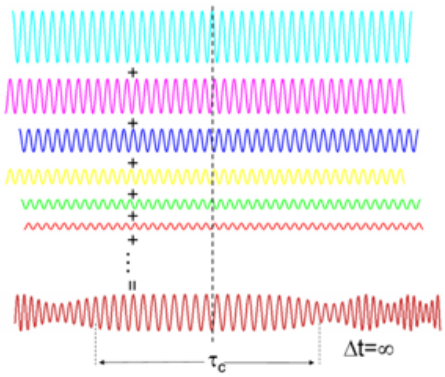


Figure 3.6: Multiple frequencies oscillating together will only have the coherence time in which they oscillate in phase.

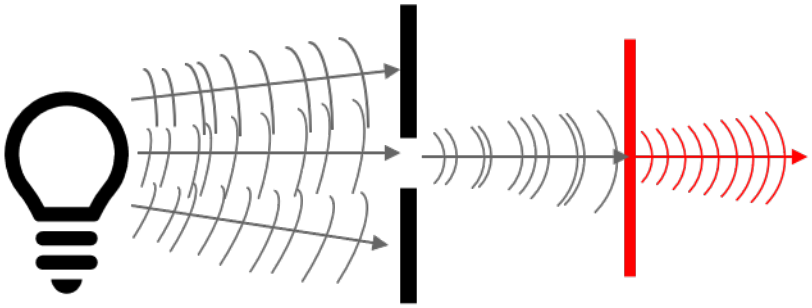


Figure 3.7: An incoherent source (light bulb) can be rendered more coherent via a spatial filter such as a pinhole and a chromatic filter.

So far, the types of coherence discussed have specifically applied to properties of a single source. If I want to compare two different sources to see how they would interfere, I need to look at mutual coherence. The most critical parameter for interference is the temporal coherence, and exactly how the different frequencies behave together strongly impacts whether two signals will interfere with each other. Thus, to measure mutual coherence, I compare the spectral densities of the signals. Spectral

densities are given by

$$S_{11}(\omega) = \mathcal{F}[E_1(\lambda)] * \mathcal{F}[E_1(\lambda)], S_{21}(\omega) = \mathcal{F}[E_1(\lambda) \times E_2(\lambda)] \tag{3.8}$$

Where $S_{11}(\omega)$ is the spectral density of signal one and $S_{21}(\omega)$ is the cross spectral density between signal one and signal two. The mutual coherence is defined based on these as

$$\gamma^2(\omega) = \frac{|S_{12}(\omega)|^2}{S_{11}(\omega)S_{22}(\omega)} \tag{3.9}$$

In essence, this compares how similar the frequency distribution of signal one is to signal two and normalizes to the total signal distribution of the two. As such, this value will range between zero and one, with one being perfect mutual coherence and zero being no mutual coherence. Of note is that only the frequency distribution is important, not the specific frequency. Thus, it is possible to have two signals separated by a great span of frequency that can still perfectly interfere, so long as the shape of the signals are identical. Thus, for my purpose I want the spectral distribution to be as similar as possible in order to obtain interference. However, that is not the only consideration when dealing with the interference of two separate sources to obtain a measurable beat frequency. For collision broadened sources (all sources that will have sufficient bandwidth such that this will be an issue), the mutual coherence function is given by

$$\gamma(\omega) = e^{-i\omega t - \frac{\tau}{t_c}} \tag{3.10}$$

which indicates that the coherence function will decrease with an increasing bandwidth. I can view this decay as being caused by the random phase walk of each source. Thus, in order to be able to actually measure a beat frequency, I will need the coherence time to be at least twice the period of the beat $t_c \geq 2T_{beat}$ to ensure

the beat is properly resolved over the phase noise. In other words, I need

$$\Delta\omega \leq \omega_b/2 \tag{3.11}$$

or $Q \leq \frac{2\omega}{\omega_b}$ for a system containing resonance. As such, the bandwidth plays a major role in whether the signals can interact or not. I find this to be the minimum requirement for interference of two signals with no definite phase relationship: the signals must have a similar spectral density so as to ensure the interference fringes have visibility, the coherence length must be sufficiently long so as to allow for detection by the diode, and the bandwidth must be narrow enough that the beat frequency can be measured within the coherence length. Thus, the entire makeup of the device depends on whether the device is intended to detect coherent radiation or incoherent, as the restricting conditions just mentioned apply primarily to the latter. Therefore I will separate discussion of device considerations into the case for coherent detectors and incoherent detectors.

3.2.1 Coherent Detector

For the case of the coherent detector, I will assume that the incident radiation will already be narrow band and thus have a long coherence time, be able to interact, and the beat be measurable above the phase noise. This substantially widens the design space, so let us begin by considering materials for plasmonic applications. As noted earlier in figure 3.4, the lower the magnitude permittivity, the stronger the field in the metal is. By extension, the larger the field in the metal, the greater number of electrons accelerated resulting in higher currents. As it happens, this also tends to be the region with the lowest loss tangent in most materials, increasing the possible signal. For the majority of materials I am interested in, a negative permittivity comes from being below the plasma frequency, so I want to operate near the plasma

frequency and just below it. The plasma frequency is given by

$$\omega_p = \sqrt{\frac{Ne^2}{\epsilon_o m_e}} \tag{3.12}$$

where N is the number of electrons, e is the elementary electron charge, and m_e is the electron effective mass. With this in mind, the easiest way to control the plasma frequency is by controlling the number of electrons. The majority of traditional metals have plasma frequencies that range from the ultraviolet to the red or near infrared. However, various semiconductors and transparent conducting oxides can be used as well, especially considering that it is easier to control the plasma frequencies in these materials by changing properties such as percents of various materials and doping concentrations. Figure 3.8 shows a few graphs of material properties that could be usable in the near infrared and the mid infrared. The choice of material, therefore, is dependent on the the designed operating frequency. The strong dispersion also contributes to selectivity in detection, especially considering the loss tangent increases the further you get from the plasma frequency.

Once the material has been chosen, the plasmon generation methodology can be determined. There are many different ways to generate plasmons on a surface, but since I am not particularly in need of concerning myself with frequency selectivity, I can consider the simplest of plasmonic generation schemes. A few of these options are shown in figure 3.9. While each of these options has their advantages, the simple groove properly tuned to the desired frequency is the simplest method for coupling from a manufacturing standpoint. An array of these grooves may also prove useful depending on the methodology chosen for implementing rectification. The most efficient coupling of these choices would be a periodic grating, either sinusoidal grating or a periodic grating of grooves.

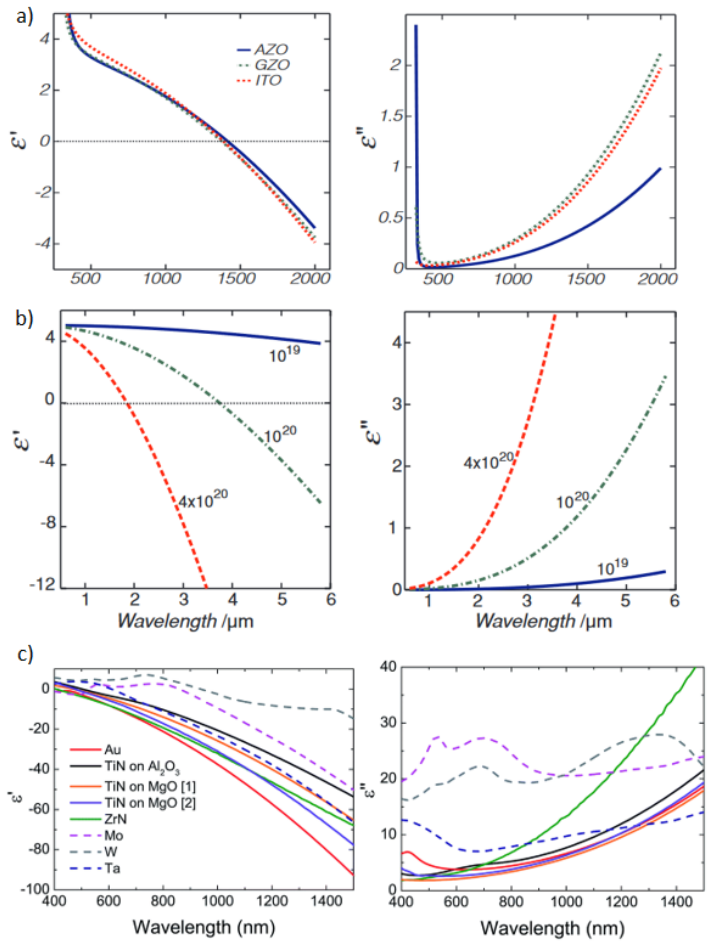


Figure 3.8: Material property plots. a) Transparent conducting oxides, with plasma frequency in the near infrared. b) Theoretical permittivity of GaN as a function of doping density [31]. Of note, any doping at or above 10^{19} is considered degenerate and can contribute to material defects. c) Permittivity of various metals [32]

After the material and the plasmon launching mechanism have been chosen, a diode of some type must be implemented to rectify the beat frequency. Naturally, the higher frequency of operation of the diode the easier it will be to design a device to operate at high frequency. This is due to the fact that as the frequency increases, the frequency separation requirement to generate a beat does not change. Since maintaining such a narrow bandwidth at high frequency can be difficult, operation at higher frequency can lead to an overlap of spectra, ultimately reducing the visibility of the beat. By increasing the frequency of operation of the diodes, you

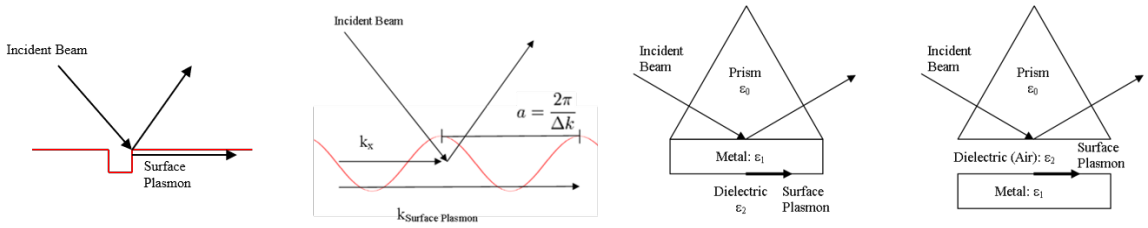


Figure 3.9: Basic methods of generating plasmons, from left to right: a groove, a periodic grating, the Kretchmann configuration, the Otto configuration. Each of these methods are designed to give the necessary momentum transfer to launch a surface plasmon.

beat frequency to be higher, relaxing the bandwidth restrictions and thus lowering the overlap and increasing the maximum operating frequency. When speaking of high frequency diodes, there are three different possibilities. First, the "geometric diode." A relatively new diode type, this diode utilizes an asymmetry in geometry to ensure electrons flow only in one direction (figure 3.10). The responsivity and IV curves for DC can be seen in figure 3.11. This diode, while it has a reasonable response, would be difficult to fabricate on a large scale. However it does have certain appeal with respect to funneling the surface plasmon waves to be rectified, although there would be additional loss due to re-emission at the edges of the diode. The second diode

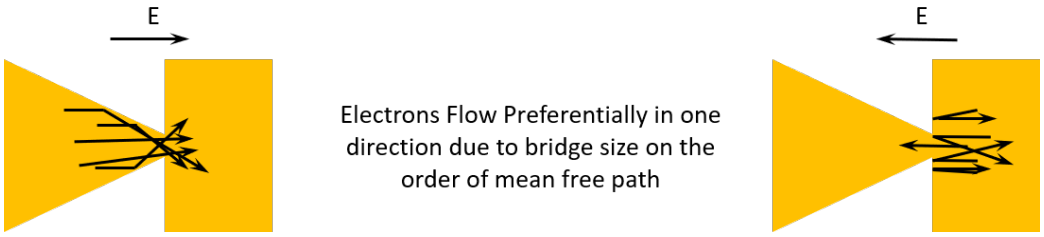


Figure 3.10: Operation principle of a geometric diode. The functionality comes from ballistic electron transport across the narrow gap.

to consider is the metal insulator metal (MIM) diode. This diode type consists of a layer of metal, a thin layer of dielectric, and then another layer of different metal. The idea is to enable electron tunneling across an uneven barrier (figure 3.12) This form of diode has been shown to rectify frequencies into the visible, although at low

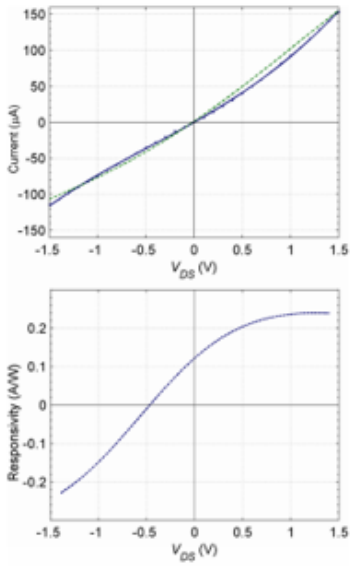


Figure 3.11: Geometric diode IV curve and responsivity curve at DC. According to [33] this diode maintains these specifications up to aprroximately 45GHz.

efficiency [34]. These diodes, while they certainly have an advantage in frequency response, are more sensitive to material choices. Since one of the "metals" is fixed from the required plasmonic response, there can be design challenges when attempting to incorporate a MIM diode into the detector schematic. The third type of diode

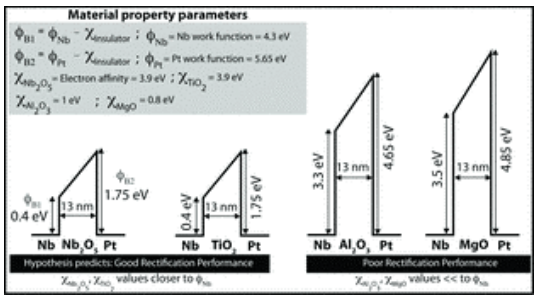


Figure 3.12: Schematic of NbPt MIM diode barriers with various dielectric layers. The asymmetric barrier heights enable rectification since tunneling is exponential. Thus, it is more likely for an electron to tunnel from the low barrier side to the high than vice versa. A high potential barrier in between the metals also inhibits rectification. [35]

to consider is the Schottky barrier diode. This diode is a standard in industry and is simply a metal-semiconductor junction [36]. This diode can easily rectify in the GHz region, and certain configurations have been shown to rectify into the low THz [37].

As one of the most well studied and implemented diodes the industrial support lies behind choosing this diode type for most design purposes, whether in planar design such that it is CMOS compatible, or in a bridge configuration for higher frequency response (figure 3.13). Each of these diodes is an applicable choice for design of a

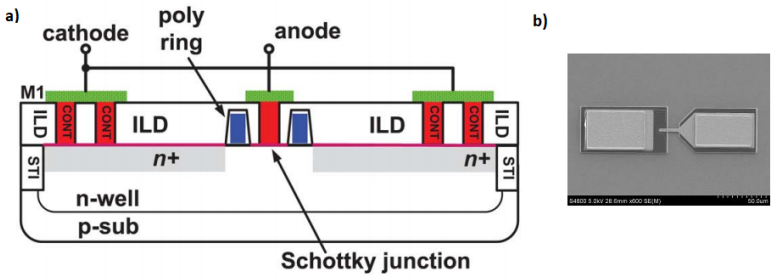


Figure 3.13: a) Schematic of a CMOS compatible Schottky diode used in a THz patch antenna array [38]. b) SEM image of a Schottky diode in the bridge configuration. The air gap reduces capacitance and significantly improves operating frequency [39].

coherent detector, though some may be more advantageous than others depending on factors such as operating wavelength and material choice. All of these diodes have a non-zero responsivity at zero bias, which means that the detector can operate under no bias and still read out a signal. It is important to note at this point that these rectifying diodes, although they have Johnson-Nyquist noise which is temperature dependent, will not rectify thermal fluctuations. This means that, between the independence of the plasmonics on temperature and the inability of diodes to rectify temperature, the entire detector should have no trouble detecting an incident signal at incident wavelength regardless of the operating temperature of the device (so long as the signal intensity is above the noise limit).

Once the diode has been chosen based on desired detection criteria, all that is left is the readout methodology. This is the most challenging of tasks to resolve and can often times structure the entire device layout. It is also, however, beyond the scope of this work. As omy goal was to establish proof of concept, a finished product level design is not currently feasible at this stage. Although some of the readout is

defined by the chosen diode, the rest can be developed further in later research.

3.2.2 Incoherent Detector

Given the problems inherent in attempting to detect incoherent radiation by any means involving interference, the simplest approach towards designing an incoherent detector is to utilize prefiltering. As shown in figure 3.7 and equation 3.7, reducing bandwidth improves temporal coherence, which is what I need to obtain a measurable beat frequency via equation 3.10. There are many methods of prefiltering, but since the goal is to obtain two narrow band peaks with minimal overlap as adjacent to each other in frequency as possible, many standard approaches may not be feasible. The two options for prefiltering are a double notch filter, which is hard to find, or two reflective bandstop filters to reflect the narrow bands onto the detector plane. While both of these options can work, they limit the entire detector plane to near monochromatic. While this can be useful, the ideal would be to integrate the frequency filtering into the device. From a practical perspective, this can best be integrated into the plasmon launching phase of design. Aside from this change, the factors in determining device operation are identical to the coherent case, as the goal here is to improve the coherence of the signal at each pixel.

There are three different possibilities for an integrated frequency-selective plasmon launching schema. The first is the case of a sinusoidal grating. The grating has a very specific period that will match the surface plasmon generation condition for a single specific wavelength (figure 3.9). Thermal noise and manufacturing defects will provide broadening to this selectivity, but it will still provide a narrow band response. In order to provide the two frequencies needed to beat, two gratings would be superimposed with the periods appropriate to the two wavelengths of interest. This would be quite challenging to manufacture and would be incompatible with most traditional fabrication techniques. Thus I will rule out this option for the purposes of this work.

The second option is to utilize slot antennae to launch the plasmons. As is well established by Babinet's principle, an antenna from a slot will behave as a dipole antenna but with an opposite polarization and impedance proportional to the inverse of the dipole impedance. Thus, if I design a slot antenna such that the resonance is at the desired frequency, the corners of the slot will efficiently couple the incident field into plasmons. For a traditional dipole antenna, the resonance is achieved when the reactance of the antenna goes to zero. Since the impedance of the slot antenna is given by $4Z_s = \eta_o^2 Z_d$, the resonance will occur at the same dimensional factor as the dipole. I can achieve resonance then by matching the length of the slot with the speed of the wave in the conductor. For low frequency radiation, this is fairly straightforward, as the velocity of the wave depends on the physical parameters of the wires used to form the dipole. For higher frequency radiation the wave velocity depends on the material properties. Whilst it may be tempting to simply utilize the refractive index, at high frequency the refractive index for metals is positive and typically much less than one. This would result in a wave with distinctly nonphysical parameters. I can, however, use my existing description of the wavevector (equation 1.1) to establish an effective wave velocity. If I use the traditional definition of wavevector

$$k = \frac{2n\pi}{\lambda} = \frac{2c\pi}{v\lambda} = \frac{c}{v}k_o \tag{3.13}$$

for a wave with (phase) velocity v, then I can compare this with equation 1.1 to obtain the fractional wave speed to scale the dipole dimension. Doing so yields

$$\frac{v}{c} = \sqrt{\frac{\epsilon_m + \epsilon_d}{\epsilon_m \epsilon_d}} \tag{3.14}$$

Thus I can expect the resonance to be somewhere around

$$l = \frac{\lambda}{2}\sqrt{\frac{\epsilon_m + \epsilon_d}{\epsilon_m \epsilon_d}} \tag{3.15}$$

This is only a first order approximation however, as other things go into the impedance of the dipole when dealing with thin metallic sheets at high frequency. The best approach is to use a full wave numerical solver to obtain the exact dimensions of the desired resonance and also determine the bandwidth of the resonance. In order to achieve the two frequencies required for the beat, I can utilize two different slot antenna designed to match the two frequencies of interest, although this may not be required in all cases as shown by simluations later. By tuning the depth of the slit correctly (via simulation optimization) such that the bottom surface acts as a back plane, I can improve coupling.

The third option to obtain frequency selectivity is through plasmonic waveguide resonators. These waveguide resonators are formed via a layer of metal, a layer of dielectric, and another layer of metal. The field couples into the resonator via the corners of the waveguide (see figure 3.2) and then will resonate within the cavity. If the bottom layer is a continuous surface, then some of the field that leaks out will escape as a traveling surface plasmon. This can be viewed as a Fabry-Perot cavity with the coupling efficiency as the input and the impedance with the air at both ends as the mirrors. I know from such a cavity that the resonance forms when the cavity is an odd integer multiple of a half wave, or in this case

$$w = \frac{(2m-1)\lambda_{sp}}{2} = \frac{(2m-1)\lambda_o}{2}\sqrt{\frac{\epsilon_m + \epsilon_d}{\epsilon_m \epsilon_d}} \tag{3.16}$$

for an integer m and waveguide width w (figure 3.14). However, knowing where the resonance happens isn't sufficient for my needs. I also must ensure the resonance is sufficiently narrow band for the interference to be measurable. To get an idea of how to approach this, I can look at an analytical model for the impedance of a plasmonic

waveguide/air interface. According to [40], this is given by

$$\eta = \frac{-2ik_{sp}}{\omega k_d \epsilon_d w} \tanh\left(\frac{ik_d a}{2}\right) \tag{3.17}$$

where l and a are dimensions according to figure 3.14. I can then use the impedance

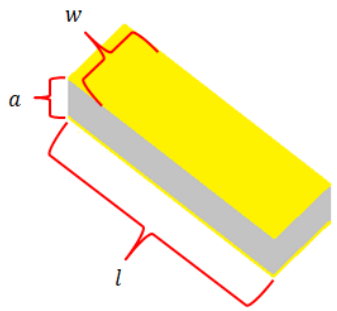


Figure 3.14: Simple diagram of a plasmonic waveguide resonator for reference with equations 3.16 and 3.17.

to obtain the reflection and transmission coefficients given by

$$r = \frac{\eta - 1}{\eta + 1}, t = \frac{2\eta}{\eta + 1} \tag{3.18}$$

which can then be inserted into the traditional Fabry Perot equations for a lossy cavity to obtain the transmission,

$$T = \frac{|t^2|}{|1 - r^2 e^{-2\alpha w}|} \tag{3.19}$$

where α is the absorption coefficient given by the imaginary part of k_{sp}; the finesse,

$$F = \frac{\pi\sqrt{|r^2|e^{-2\alpha w}}}{1 - \sqrt{|r^2|e^{-2\alpha w}}} \tag{3.20}$$

and the free spectral range.

$$\nu_F = \frac{c}{2w}\sqrt{\frac{\epsilon_m + \epsilon_d}{\epsilon_m \epsilon_d}} \tag{3.21}$$

which may be noticed is simply the frequency separation between resonances as given in equation 3.16. From here I can find the line-width of the resonance by dividing the free spectral range by the finesse $\Delta\nu = \nu_F/F$. Thus, I can know whether the bandwidth is sufficient, and if I also know the coupling efficiency to the resonator I can establish the coupling to the energy in the plasmon modes. In addition, I can use this as a starting point and obtain numerical results using computational approaches.

I have established methodologies to obtain narrow bandwidth for the incoherent case, but now it becomes necessary to determine the detectable power from an arbitrary incoherent signal. Equation 3.6 is the maximum rectifiable power from a coherent signal, so this is a good place to start. All that is required to modify this to the maximum detectable power for an incoherent source is to determine the relative overlap of the detector with the incident signal. To do this we can use the definition of the mutual coherence to compare the sensitivity of the detector to the incident radiation. Therefore the total rectifiable power of the device should be

$$P = P_b * \gamma_{ds}^2 \tag{3.22}$$

where γ_{ds} is the cross spectral density function of the detector sensitivity and the source. All together, these options open up the design space of the incoherent detector, whilst still being similar in many respects to the coherent case. Now I can begin designing individual cases in an attempt to obtain a working proof of concept.

CHAPTER 4: DEVICE SIMULATIONS

For practical design considerations, I first should simulate electromagnetic device performance. Ideally an integrated physics model could be utilized to include diode performance, but in practice this can be tricky for full simulations of new devices. As such, I will focus on the electromagnetic properties of the devices and look to established diode responses for functionality past that point. To begin with, let us take a look at two of the plasmon coupling methods. First, I will examine the plasmonic resonator launcher as shown in figure 3.14.Several simulations have been carried out regarding the coupling efficiency and overall effectiveness of these resonators. First, I began with 2D simulations to determine the efficiency of coupling to an infinitely long rectangular waveguide. The geometry for these simulations is shown in figure 4.1. For this simulations, I chose to utilize a full field simulation, as utilizing a scattered

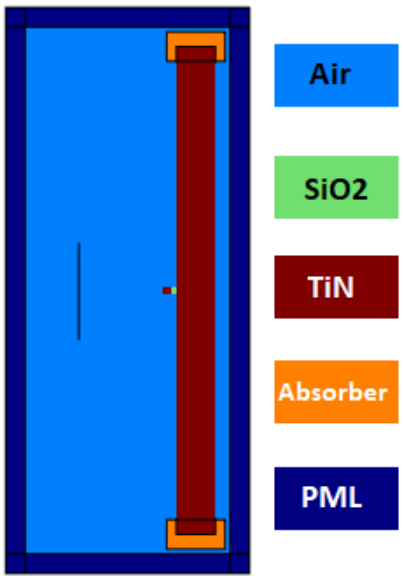


Figure 4.1: Geometry and materials of the simulation for the plasmonic waveguide resonator. The absorbing end caps are to prevent interference from scattering and reflection at the edges.

field approach can result in a difficulty visualizing the plasmons without a muti-step process. Many approaches for incident field were attempted, but in the end I utilized an internal port to feed a plane wave into the region of interest. This limits the interference between the incident field and the scattered field by limiting the illumination range. Whilst this could also be accomplished by a structured background field, such as a Gaussian, the change in intensity along the spatial profile resulted in skewed results for coupling efficiency. By utilizing a port with the slit condition activated on the internal boundary, I was able to generate a plane wave solely in the area of interest; and by making the slit sufficiently large with respect to the wavelength, I was able to minimize diffraction effects. I then established a region of integration a distance sufficiently far away from the resonant element so as to substantially reduce any signal from scattered waves that were not surface coupled. Since the resonator back-scatter resembles a $cos^2(\theta)$, this distance was not so far such that losses became a significant problem. A line segment was selected along this region whereupon the current density norm was integrated. By integrating the current density along this region it is possible to obtain at least a qualitative comparison of coupling efficiency, since a higher coupling efficiency will result in stronger surface plasmons which will result in a higher current density.

In order to turn this measurement into a quantitative comparison, the field at the surface will have to be compared to the incident field and expressed as a ratio (equation 4.1). The integrated current density was then used as an objective function utilizing the optimization module in COMSOL to optimize the physical geometry of the plasmonic waveguide resonator. The starting point for this optimization is given as the theoretical ideal from equation 3.16. The adjustable parameters for the two dimensional case were the width, dielectric thickness, and top metal layer thickness. For 2.2 micron radiation incident upon a material of permittivity $\epsilon = -5 - 0.5i$, the ideal dimensions for a resonator were $w = 0.534\mu m, t_d = 0.408\mu m, t_m = 0.815\mu m$.

The full simulation can be seen in figure 4.2 and a close up of the launched plasmons is shown in figure 4.3. Utilizing these dimensions, I then swept over frequency to

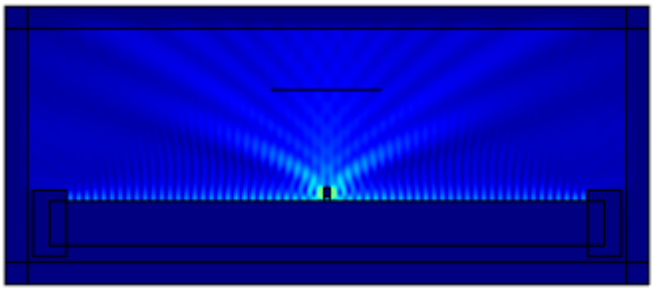


Figure 4.2: Full simulation of the plasmonic waveguide resonator. Of interest is that the optimal coupling occurs when the top metal layer is sized to produce it's own resonance from the surface plasmon waves, increasing the coupling to the waveguide resonator.

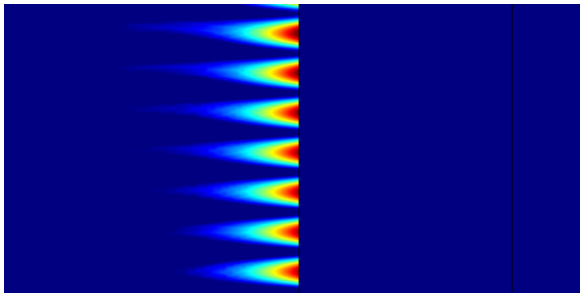


Figure 4.3: Closeup of plasmons launched from a plasmonic waveguide resonator. This simulation result clearly shows the surface waves and the field in the metal.

determine the bandwidth of the response. This result is shown in figure 4.4. Based on an initial field amplitude of 1 V/m, the coupling efficiency is then 68%, according to the formula

$$\xi = \frac{\overline{|E_{spp}(\omega)|}}{\overline{|E_o(\omega)|}} \tag{4.1}$$

where the bar indicates a time average, and the surface plasmon field is taken at a point close to the source. This is simply comparing the average power for the simulated surface plasmon to the ideal case where the incident power is completely coupled to the surface plasmon. In the time averaged case, this value should have a maximum of k_{spp}/k_o and a minimum of zero, although practically speaking it will

likely always be less than unity for the cases I am working with.

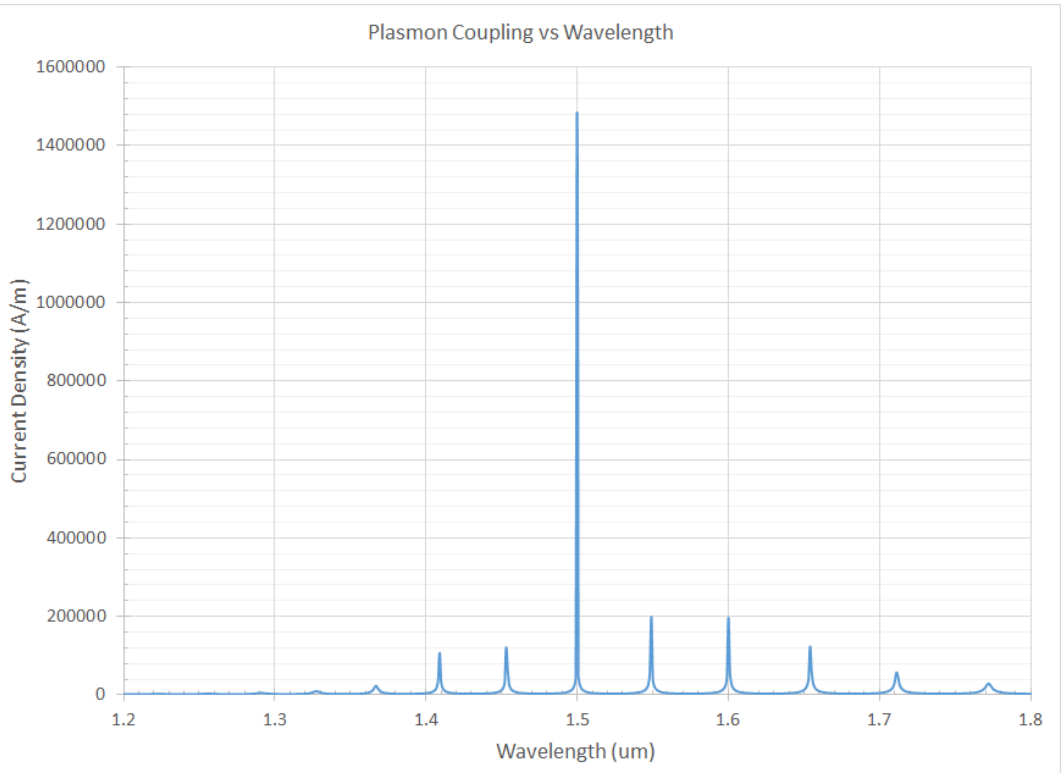


Figure 4.4: Bandwidth of the plasmonic waveguide resonator for illumination at 1.5 microns. Two important things can be drawn from this result. First, the Fabry-Perot type resonance peaks as predicted by equation 3.20 and 3.21 and secondly the narrow bandwidth (1nm) allows for proper frequency downselection to allow for interference to generate a beat.

Once the peak is determined, I can then choose a nearby frequency based upon the bandwidth to provide high interference visibility and repeat the process, but keeping the dielectric and metal thicknesses constant (for ease of fabrication). I can then take the results from the two different frequencies and sum them together to see what the interference pattern would look like for two separate resonators generating the fields co-linearly. An example of this with frequencies sufficiently separated so as to be easily seen is shown in figure 4.6. These results indicate that it is not only feasible to utilize plasmonic waveguide resonators to launch surface plasmons, but that the resonators can have sufficiently narrow bandwidth to function in the mid infrared. I can then use this to determine the total power efficiency for the device. From

equation 3.6 I can obtain the maximum power in the side band, so I can take the coupling efficiency multiplied by the sideband power to determine the total detectable power in the sideband as a fraction of the incident power. This is given by

$$P_{eff} = \xi P_b = \xi P_o \frac{\left(\frac{A_{max} - A_o}{A_o}\right)^2}{4} \tag{4.2}$$

Where A_{max} is the maximum interference amplitude and A_o is the incident field amplitude. For the plasmonic waveguide resonator demonstrated here, the theoretical total power efficiency is 17% of the initial power. To turn this into a responsivity for the detector, you simply multiply this value by the responsivity curve of the diode you are utilizing for rectification at the beat frequency, as the power efficiency gates the incident power. Thus, if the diode has a responsivity of 0.5 A/W, the responsivity of this detector would be 0.085 A/W.

At this point it becomes reasonable to compare the theoretical specs of a device such as this with a traditional photodiode. Since our target wavelength is 1.5 microns, it is reasonable to compare to a Germanium photodiode. The quantum efficiency of such a diode is shown in figure 4.5 with variance attributable to the specific manufacturing methodology [41] and lies between 60% and 80% at peak. This is comprable to the coupling efficiency of this device. Although the photodiode has a broader bandwidth, it does not inherently provide frequency selectivity. A filter is typically used in conjunction with a photodiode to retrieve frequency information, so the specific filter used would reduce the photodiode efficiency further while simultaneously narrowing bandwidth. A typical notch filter, such as those found on ThorLabs, would have around a 45% transmission and a 12nm bandwidth. Combined then, this would provide a Ge photodiode desiring frequency selectivity at 1.5 microns an efficiency of approximately 31.5%. Our detector, at 17% is a little over half as efficient, which for a first round of simulation and development is a reasonable achievement. If the

additional benefit of being able to determine relative phase information is added to this, the difference in performance is a reasonable exchange even if later engineering improvements on device efficiency were not assumed.

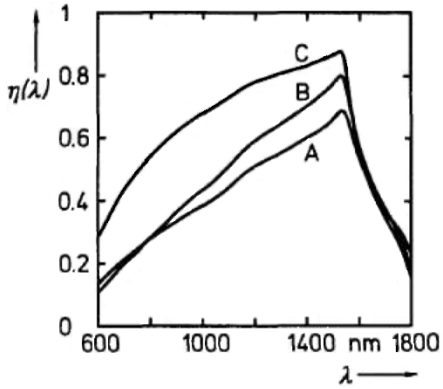


Figure 4.5: Quantum efficiency of three different Ge photodiodes [41].

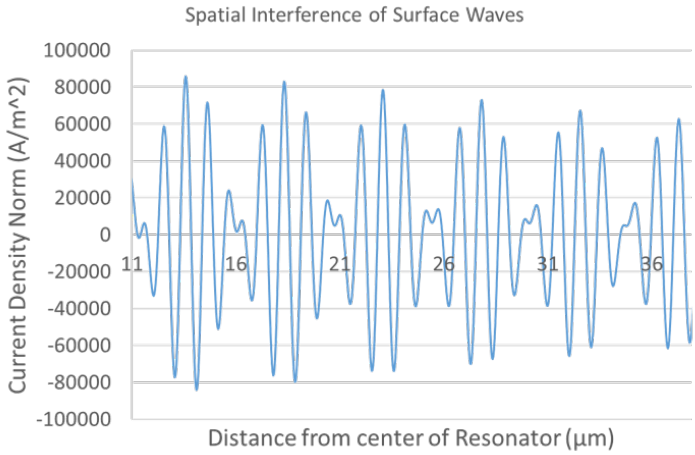


Figure 4.6: Interference between two different plasmonic resonators operating at two different single frequencies generating a spatial beat.

These studies were done at normal incidence and single frequency. In order to get an idea of what to expect from broadband illumination or non-normal incidence, I need to vary the simulation some. If I wish to utilize this structure as a detector, the angular field of view needs to be assessed. In order to do this, I can vary the angle of incidence of the plane wave by some angle. Figure 4.7 shows an angular half sweep from 0 to 45 degrees, with the current density evaluated in the same manner as

before. Normal incidence coupling is high, and a 11.4 degree incidence couples even

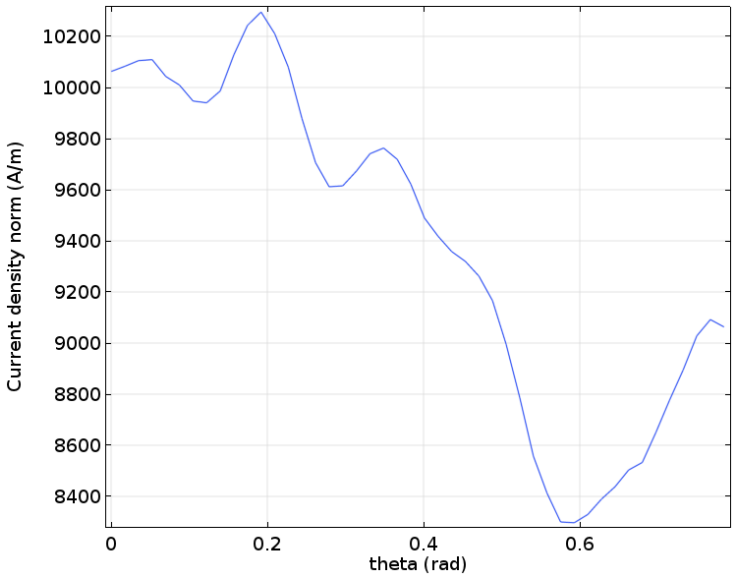


Figure 4.7: Angular response of the plasmonic wavguide resonator. Plasmon generation is most efficient at 11.4 degrees, falling to 80% generation efficiency at 34.4 degrees.

better to the surface waves than at normal. There is however, a region where the generation efficiency drops to 80% of maximum at 34.4 degrees.

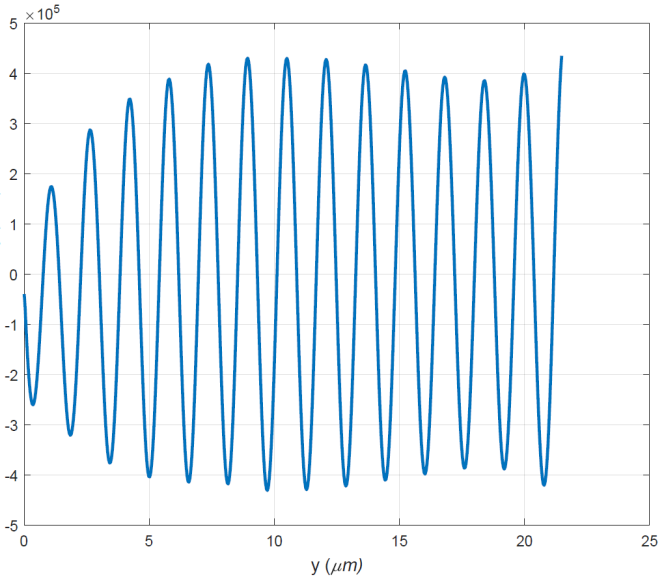
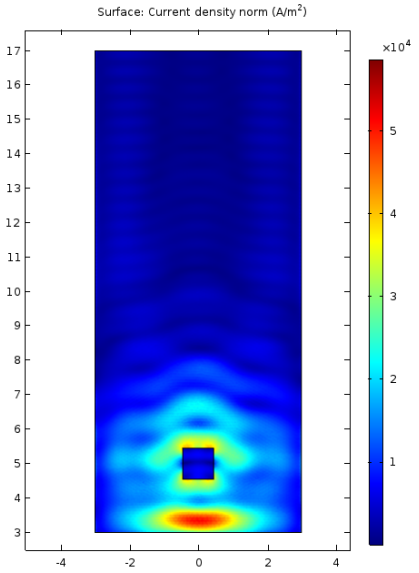
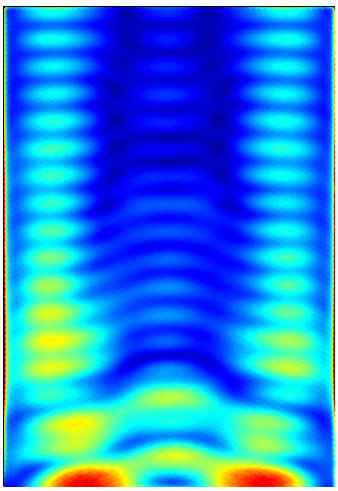


Figure 4.8: Broadband illumination result from simulating the bandwidth from figure 4.4 as incident. Although there is a slight variance in amplitude, the fundamental frequency is still quite clear, indicating that the resonator will perform well under broadband illumination.

This can be calibrated for, although it might be compensated for ore completely avoided when designing a focal plane array for the device. In order to simulate a broadband illumination, I summed together the solutions for a gaussian distribution of frequencies.



(a) Unoptimized Resonator in 3D



(b) Optimized Closeup

Figure 4.9: Simulation of a square plasmonic waveguide resonator in three dimensions. The plane shown is the surface of the metal.(a) Simulation of resonator at same physical dimensions as in 2D. (b) Close up of re-optimized resonator field. Truncating the resonator in the third dimension shifts the resonance slightly, as would be expected.

Since COMSOL tends to prefer single frequency input, and I can choose materials for which there is minimal nonlinearity at the powers I am utilizing, simulating multiple single frequencies and then summing them together is a perfectly reasonable approach to obtaining a broadband solution. So long as the frequencies are relatively close together, they will accurately approximate broadband illumination. The results of this simulation can be seen in figure 4.8. It is fairly clear to see that even in the event of broadband illumination, a signal can still be obtained. Following all of this, simulations were done in three dimensions to confirm the functionality of the resonators when truncated in two directions. Figure 4.9 shows the simulation results for this arrangement.

The plasmonic waveguide resonator is only one of the better ways to couple surface plasmons while incorporating frequency selectivity. I still need to examine the ideal slit antenna coupling properties.

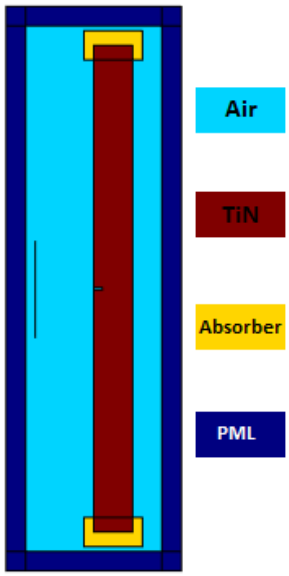


Figure 4.10: Geometry and materials of the slit antenna simulation. The line in the center left is the excitation port. The impedance matched end caps are designed to reduce scattering and reflections off of the end of the metal bar.

In order to accomplish this, I kept the same basic geometry as the simulations for the waveguide resonator, but replaced the resonator with a slit (figure 4.10).

The integral to determine current density is still taken in the same manner, as it proves to be an effective methodology for determining coupling efficiency. I still utilize the integrated current density norm as an input for the optimization module in order to determine the ideal geometry. My starting point for this optimization is the ideal from equation 3.15. For these simulations, my illumination wavelength was 1.5 microns and the metal was crystalline Titanium Nitride [42]. These choices are pertinent to the next chapter. The determined ideal width and depth of slit were $w = 0.2655\mu m$ and $d = 0.8892\mu m$. These parameters correspond to a coupling efficiency of $\xi = 10.5\% \pm 1\%$, where the uncertainty is due to the proximity to the slit relative to the absorption losses at the distance the measurement was taken. The full simulation for this is shown in figure 4.11. After the full simulation was done, I

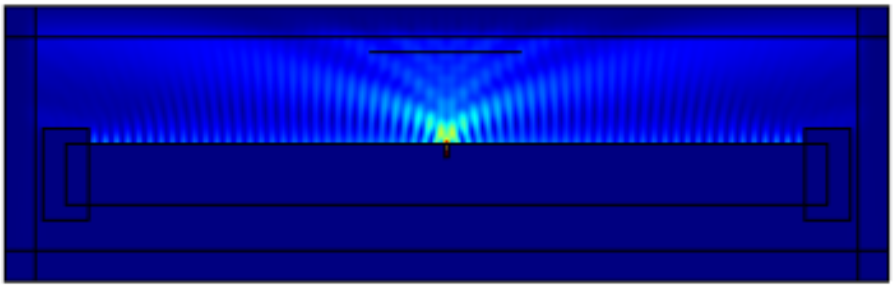


Figure 4.11: Simulation for the slit antenna at the optimum geometry.

swept the wavelength to determine the broadband response of the slit. The results of this sweep are shown in figure 4.12. As is clear, there is a resonance at 1.5 microns and another at 1.47 microns. This is actually a good combination, as it provides a beat frequency of 4.078 THz, or a wavelength of 73.5 microns, which is in the realm of detectability by a bridge Schottky diode or a mim diode at reasonable efficiency. I then followed the same approach as with the plasmonic waveguide resonator and interfered the two peak frequencies. Although the simulation is not of sufficient size (read memory) to see even one period of the beat, by adjusting the phase of one signal with respect to the other, I can find the maximum and minimum amplitude of the interference pattern. This gives us the visibility of the fringes and allows us to

calculate the effective usable power from the signal. In this case, I have a visibility of 68.5%, which means that for this particular arrangement, 1.2% of the incident power will be rectifiable by the device. How much current output that will give depends on the responsivity of the diode used in rectification which was not included in the simulation. This establishes the slit as effective for use in both a coherent type and an incoherent type detector, as the bandwidth of each resonance was approximately 2 nm or 271.9 GHz. This gives, according to equation 3.11, plenty of room to work with the signal without fear of a random phase walk destroying the relative phase information. As a further step to verify that this would function as an incoherent

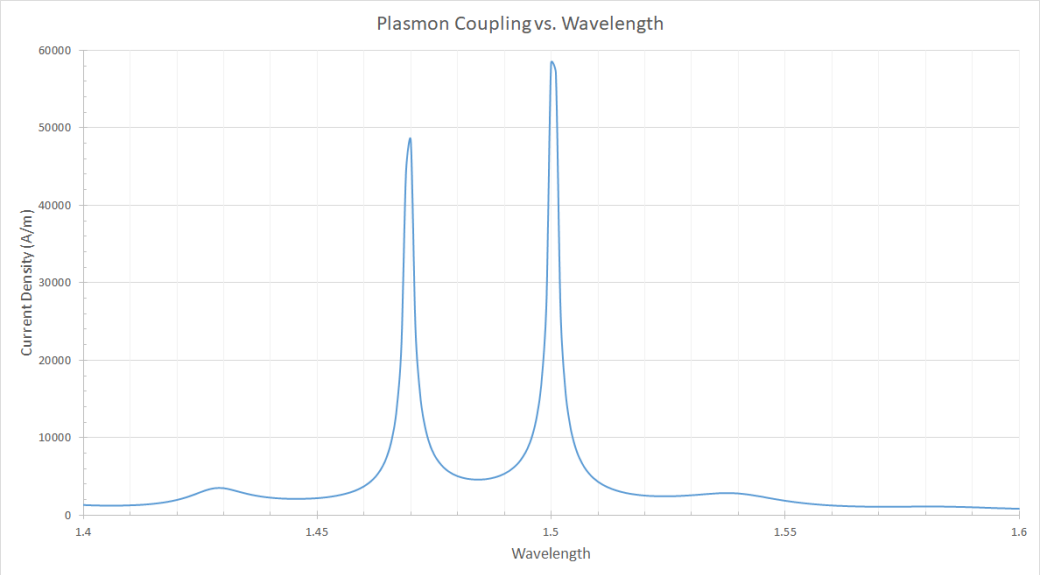


Figure 4.12: Simulation results showing the integrated plasmon intensity for a single slit antenna for a range of wavelengths.

detector, I took the sum of all of the wavelengths swept in figure 4.12 to represent a broadband illumination (500nm bandwidth, which is sufficient as there are no other significant resonances for this feature anywhere near these two peaks). The result is shown in figure 4.13. Again, the simulation was not sufficiently large to clearly see a full period of the modulation, but through modulating the phase it was clear that there is modulation to the overall interference and that the oscillation is not washed out over the entire length of the simulation. With the relative incident phase

adjustment providing the maximum and minimum of the peaks, the slit provides a 47% modulation, a decrease of only 11.5% from the ideal two frequency case.

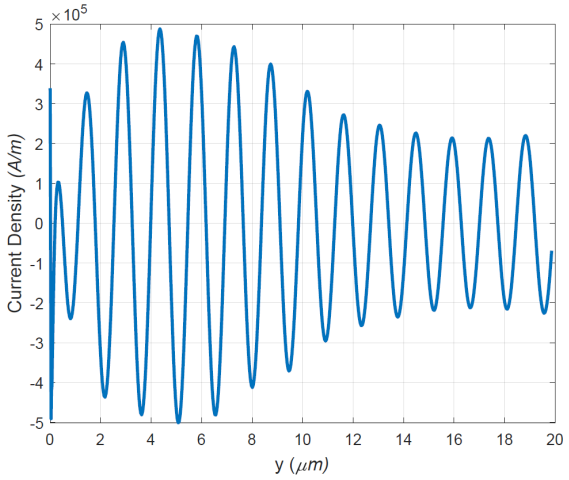


Figure 4.13: Simulation results showing the summed field over the 100nm bandwidth from figure 4.12.

With the slit antenna validated as a choice for plasmonic coupling in both coherent and incoherent arrangements, I now look to an angular study to verify the functionality for use in a focal plane array. As in the plasmonic waveguide resonator, I swept the angle of incidence of the field and looked at the integrated current density to determine the coupling efficiency. The results from the sweep can be seen in figure 4.14. The coupling remains high even at high angle of incidence, only dropping by 10% at 45 degrees, and even returning to optimal coupling at 30 degrees.

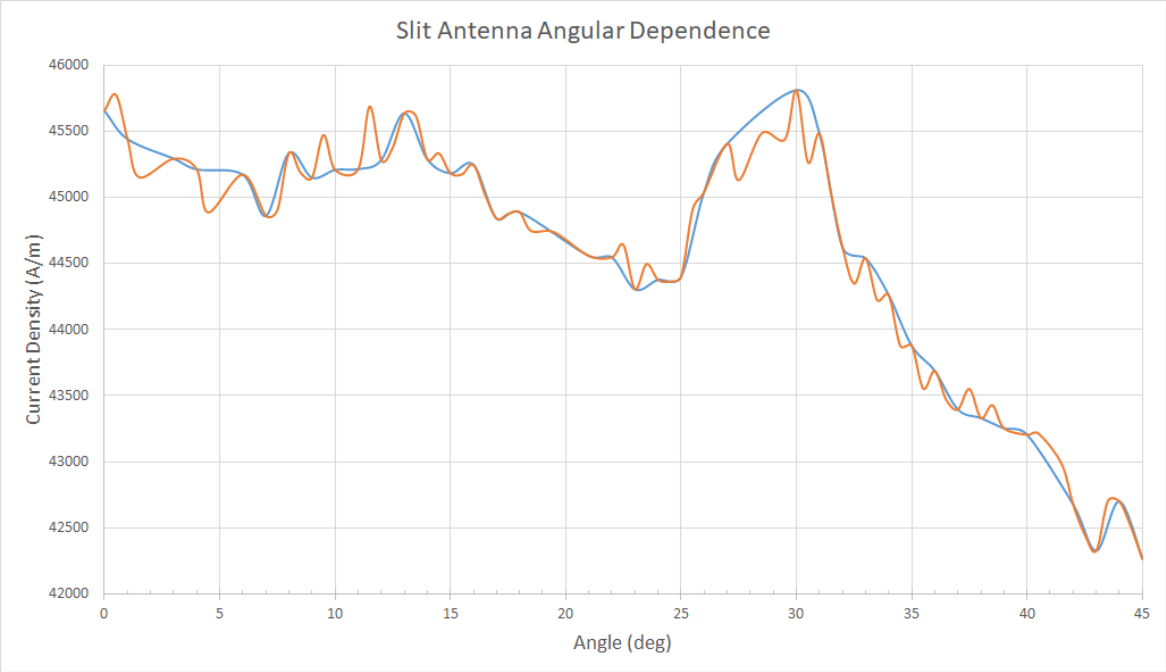


Figure 4.14: Simulation results showing the angular coupling from integrated current density due to launched plasmons. This was obtained by integrating the current density over a region whilst sweeping the incident angle of radiation.

CHAPTER 5: FABRICATION

First, let us discuss some general fabrication methodologies as they apply to a functional implementation of this device. The first step in fabricating these devices is choice of material. Whilst simple metals and readily available transparent conducting oxides can generally bridge optical and NIR, in order to go to MIR or LWIR an intentionally doped semiconductor must be used [31, 43, 44, 45]. The doping concentration must vary according to where the desired plasma frequency will be located. This means fabrication must begin with deposition chambers. Since the goal is to operate near the plasma frequency, the material need not be perfect; all that is required is that the permittivity is negative and the loss fairly reasonable. The thickness of the material and whether multi-step deposition is required depends on the coupling mechanism. If utilizing a plasmonic waveguide resonator, then a multistep process must be used. Fortunately, the semiconductors used for doping are effective dielectrics if they remain undoped (or unintentionally doped) for all frequencies this approach would be applied to. This simplifies the multistep process into a single growth with controlled doping. The material choice can be made somewhat more difficult by the requirement of inclusion of a diode. The main contenders for industrial applicability in design are the Schottky diode and the metal insulator metal diode. Both of these diodes require engineering the bands of a semiconductor relative to a metal. This means that the diode must be integrated in to the choice of metal for the device. This integration means that further design is needed for many device designs operating in the MIR and LWIR to ensure that the diode is compatible with the design process.

Once the material and diode type is chosen for the desired frequency of operation, the patterning can begin. Most patterning (virtually all industrially compatible pat-

terning) consists of creating a mask of either the pattern or its inverse, then etching away the unwanted material (figure 5.1). The mask is typically defined by a pho-

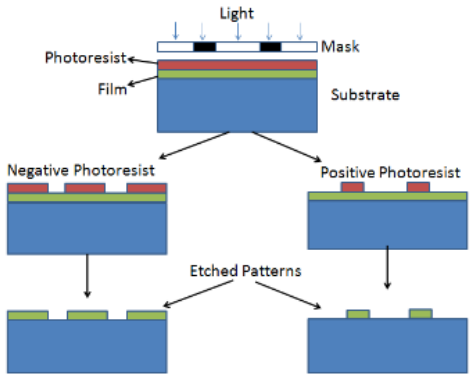


Figure 5.1: Example photoresist exposure and the resulting pattern after etching.

tolithography process in which a photomask is made which will block or phase shift ultraviolet radiation to form a pattern [46]. The pattern will, under flood illumination by ultraviolet, expose a region of photoresist. Depending on whether the resist is positive or negative, the exposed region will either polymerize or depolymerize respectively. Once developed, this pattern will form the basis for the device. At this point, the etching methodology must be decided. The two main categories of etching are physical and chemical etching. Physical etching can be subdivided into reactive and non-reactive. Reactive etching utilizes a chemical reaction with a gaseous vapor to remove exposed material. This process can be accelerated by utilizing an bias voltage to increase the velocity of the reactive gas hitting the surface. Non-reactive etching essentially utilizes a stable particle to physically dislodge material. This can be anything from an ionized noble gas such as argon, all the way to sand blasting. Physical etches tend to be less sensitive to material, whereas chemical etches tend to be very material dependent. Some chemical etches for example, are far more aggressive to photoresist than to the desired etch material. For materials like this, a hard mask must be used. This can be done either via a lift off process or a two step process. In a lift off process, a photoresist negative is put down with a negative angle sidewall

(akin to an inverted trapezoid). The hard mask is then deposited via sputtering so as

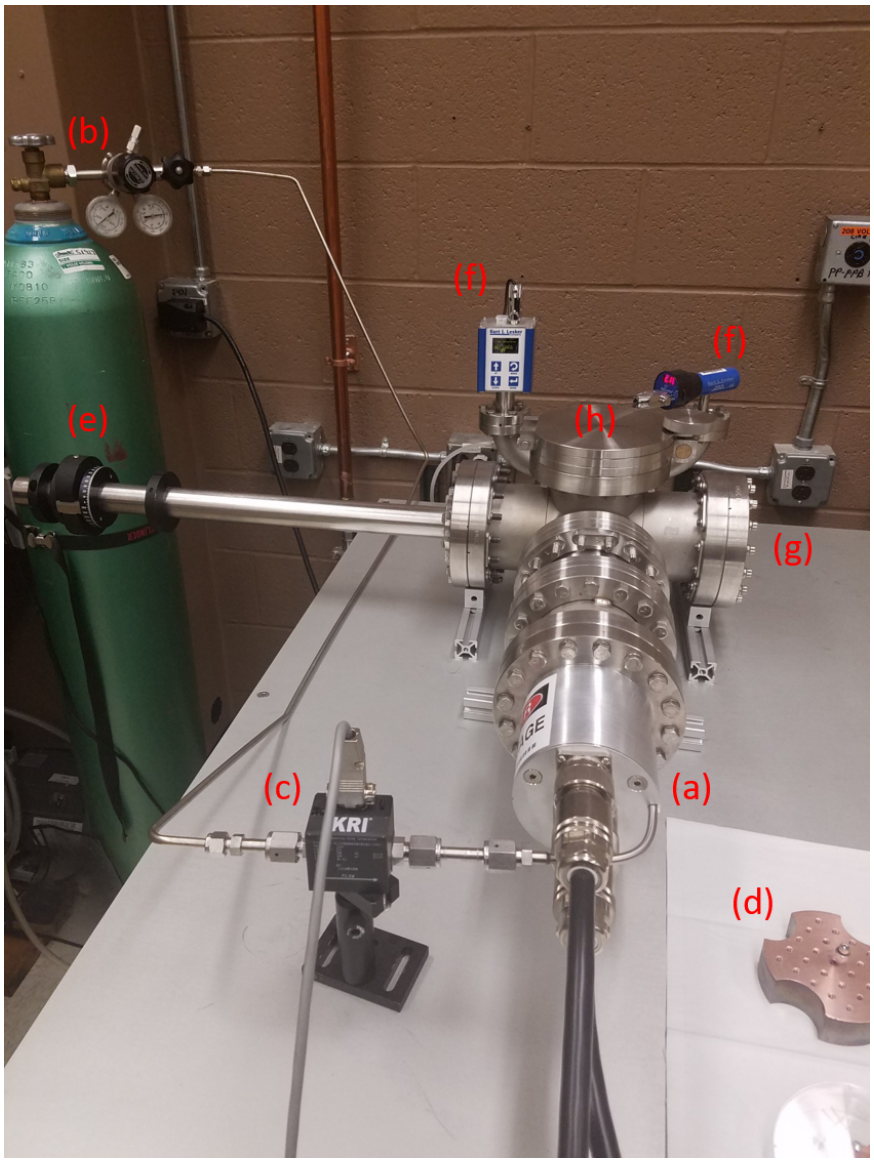
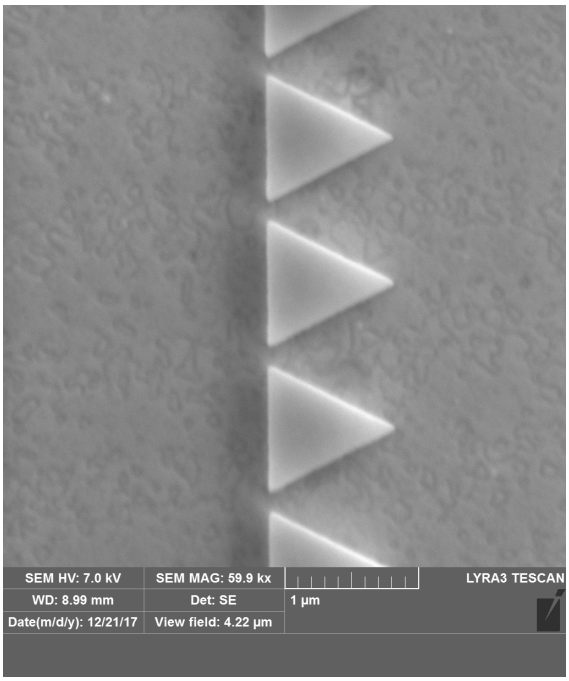
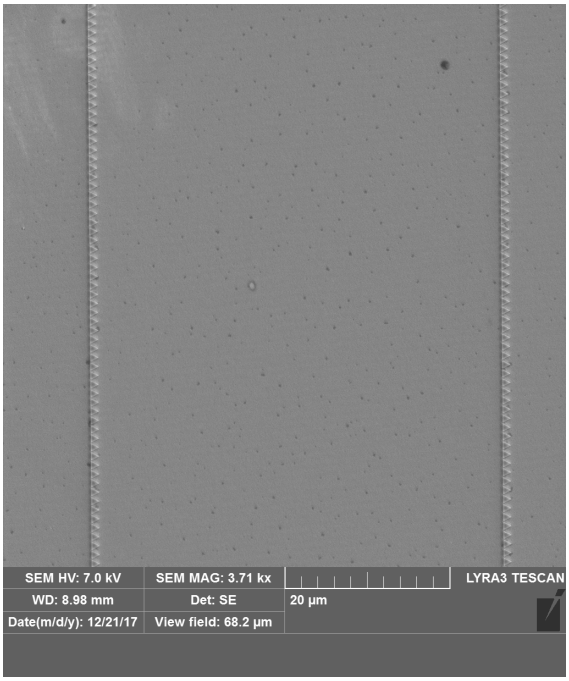


Figure 5.2: Diffuse Ion Mill image. (a) Ion gun. Possible acceleration voltage ranges between 200V and 1000V. (b) Argon gas supply and pressure regulator. (c) Mass flow controller. (d) Sample holder. (e) Beam shutter. (f) Pressure guages. (g) Observation window. (h) Flange for inserting and removing sample.

to reduce sidewall deposition. The resist is then removed (typically via ultrasound) leaving the negative mask pattern. In a two step process, the hardmask is deposited directly upon the substrate and the photoresist is patterned on top. An etch is then used to remove all of the hard mask except for the desired features. This leaves the



(a) Closeup of diodes



(b) Diode full profile

Figure 5.3: (a) Close up of the geometric diode array mask. Since the chemical etch will undercut the hard mask, the critical feature spacing had to be increased from 40nm to approximately 100nm. (b) Zoomed out view of the mask for the geometric diode array.

mask as the positive for the desired etch pattern. In either case, the hard mask is prepped for the reactive etch. These are but generalizations of broad methodologies, each of which could be dealt with at length an in far more detail, but for now it will suffice to only discuss in more detail those methodologies which apply to a device which I intend to fabricate for a proof of concept.

To fabricate a proof of concept device, I needed to choose parameters which would function well together. Although the design space is vast with many options that would eventually lead to a functioning device, I decided to pursue a coherent detector with a slit coupling. This choice had multiple driving factors. First, I needed to pursue a proof of concept that would experimentally validate theoretical predictions, for which a coherent detector would do nicely. The slit coupling via simulation in TiN worked well for both coherent and incoherent approaches, so it proves to be a viable test candidate. Of diode choices I decided to attempt the geometric diode. This was due to two main factors. First, the in plane nature of the geometric diodes would allow for direct observational analysis to ensure the fabrication was according to spec as well as utilizing the in plane field vectors without the additional complication of the exponential decay through the metal film. Secondly, the geometric diodes are reported to have a non-zero responsivity at relatively high frequencies, ensuring that the beat will be rectified. Along with this, I decided to choose TiN as the material because the constriction point of the geometric diode will experience high Joule heating which may cause less robust materials to fail with such a small dimension. For example, with the current generated from our ideal slit earlier, TiN will experience a temperature rise of approximately 1700K. Since TiN melts at 3200k, the material should not melt or deform due to the current passing through the diode. If we were to use Gold, which has similar conductive properties but a melting point of only 1300K, the material would melt. The wavelength choice will be set at 1.5 microns for two reasons. One, the TiN still has reasonable properties

in this wavelength range; two it will demonstrate the feasibility for NIR work, which is important because this device becomes more and more difficult to implement at higher frequencies due to the narrower required linewidths for good interference. Thus while these choices are perhaps not the most ideal of combinations for an end scale industrial device, this approach provides insight into the fundamental operational methodologies that will allow an industrial type redesign to operate more efficiently. Figure 5.4 shows a representative model of the experiment and detector layout.

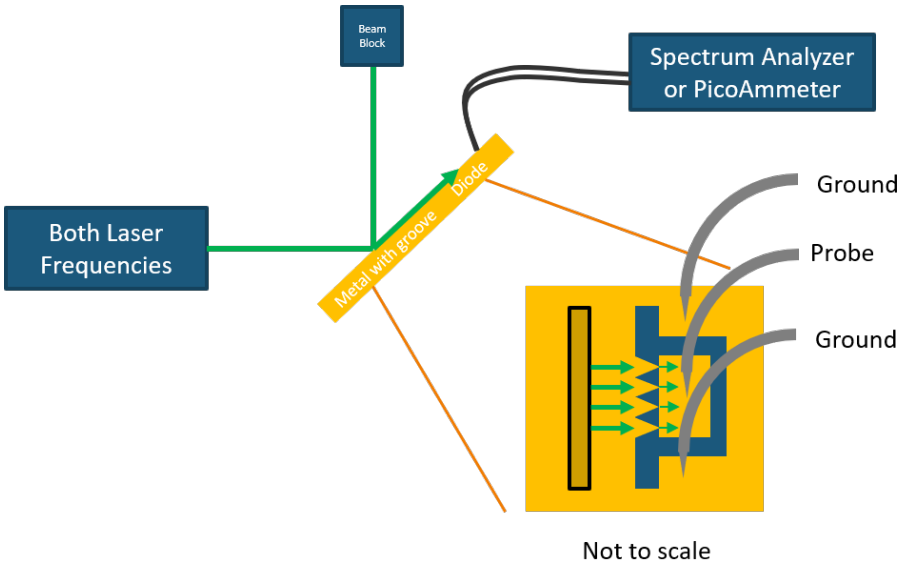


Figure 5.4: A representative model of the experiment and detector.

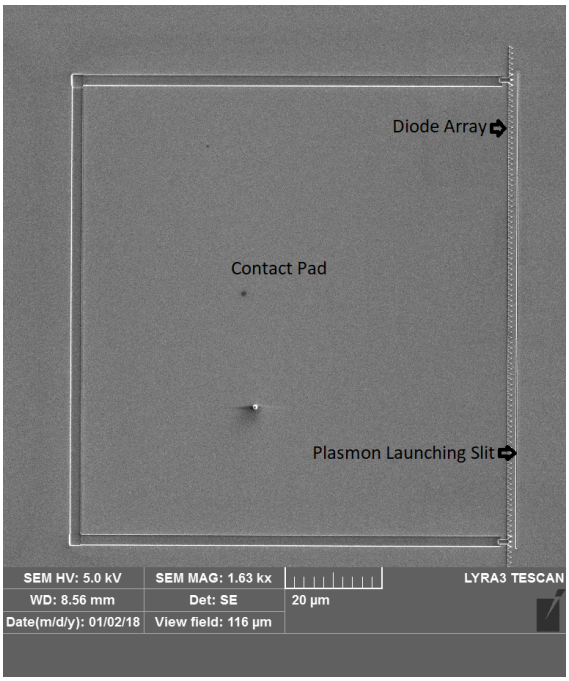
To begin, let us first examine methodologies of patterning TiN crystalline films. TiN is considered one of the harder, more chemically robust and thermally stable materials to have been discovered. It appears a golden color but is much more robust than gold. Its common uses are as drill bit coatings and for plating in costume pieces as gold replacement. Because of the stability of this material, patterning it can be somewhat challenging. I have found three methods that can reliably etch a TiN film. The first is a direct etch utilizing an accelerated ionized argon gas. Figure 5.2 shows the chamber which is used for the direct etch.

An argon gas is fed into the ion gun which, utilizing columnating molybdenum particle optics, ionizes and accelerates the gas to a 2 cm radius at the other end of the chamber. I built this chamber specifically for the fact that accelerated argon does a good job of etching most materials at a fairly constant rate. Due to the chemical stability and hardness of TiN, however, this rate is substantially reduced. After conducting several tests in which a portion of a TiN film was masked and another was left bare, I was able to determine that the etch rate of TiN from an accelerated argon gas with an 800 V acceleration potential was 9 nm/min. Comparatively, this process removes most resists at a rate of somewhere between 40 and 60nm/min. For any but the smallest of features or surface finishing, this would be difficult due to the required resist thickness.

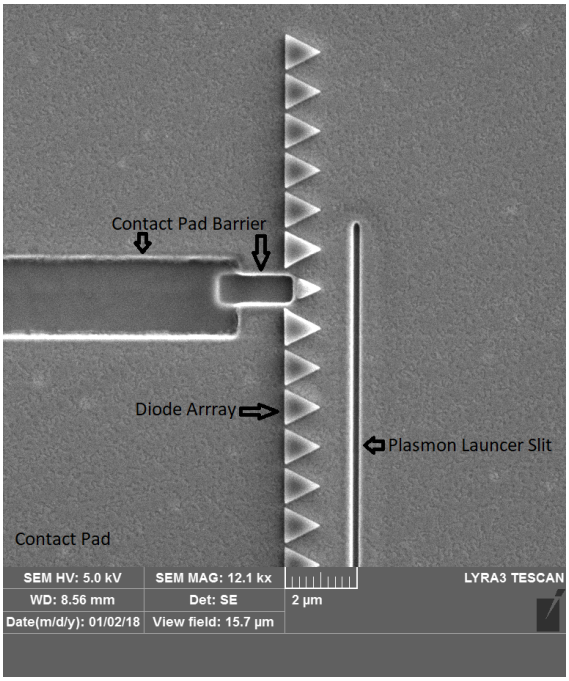
The second method of pattern transfer is a chemical etch; specifically one part ammonium hydroxide, four parts hydrogen peroxide, 12 parts water. This is traditionally referred to as an APM etch, but is more commonly used as a surface preparation. The hydrogen peroxide is a strong oxidizing agent, which attacks a surface and converts it to an oxide. The ammonium hydroxide is then dissolves the oxide away. This is a much faster process, with etch times in the 20-50 nm/min range. Whilst the one to four ratio of ammonia to peroxide seems to be critical, varying the water concentration only changes how aggressive the etch is. The only difficulty with this etch is that it is far more aggressive to polymer resist than to TiN. Thus, in order to fabricate a mask to transfer a pattern with this etch, a hard mask must be used. I found that aluminum is resistant to the APM etch, and is compatible with a lift-off methodology. I first deposit a layer of lift off resist; for the samples described later on this resist was polymethyl methacrylate (PMMA). For my work, I utilized an e-beam to write the patterns since the mean free path length (the critical dimension of the geometric diode) of TiN in around 45nm. Since my goal is image transfer, I exposed the pattern I wanted transferred, then developed the resist. This left us with the

desired features in polymer on the surface. I then sputtered aluminum (sputtering is necessary to prevent sidewall deposition) onto the sample and put the sample in ultrasound to remove the polymer and expose the TiN film in the areas I want to transfer the pattern. This allows us to proceed with the chemical etch to transfer the pattern directly into the TiN film. The resultant diode masks can be seen in figure 5.3. The third method I found to work in TiN etching is to utilize a reactive ion etching. Specifically chlorine gas, when slightly accelerated (18V acceleration) will etch TiN very well and can be used with standard resists. By utilizing traditional photolithography along with this method, large scale production is far more feasible. The facilities I used were not however, equipped to handle reactive ion etching with chlorine, so I will table that option until a later date.

With the diode masks in place, the next step is to modify the mask for the larger features that are not as dimensionally critical. Had I been able to take a photolithography approach, I could have done all of the masking in one step. However since I was using electron beam, the contact pads and plasmon launching slit would be impractical for e-beam due to the relatively large size. Thus I used a focused ion beam to pattern the contact pads and slit pattern through the hard mask. It was fairly easy to selectively etch the mask since TiN is highly resistant to FIB ablation. The complete pattern can be seen in figure 5.5 as well as a close up of the slit mask. In order to dest functionality of the diode in both directions, contact pads were made in pairs, with the pad on different sides of the diode in each pair, as can be seen in figure 5.6.



(a) Diode full profile



(b) Closeup of diodes

Figure 5.5: (a) Zoomed out view of the mask for the entire detector, including the contact pad. (b) Close up of the area where the contact pad meets the diodes and the slit. The smaller cut is to ensure proper conductive isolation between the contact pad and the rest of the metal.

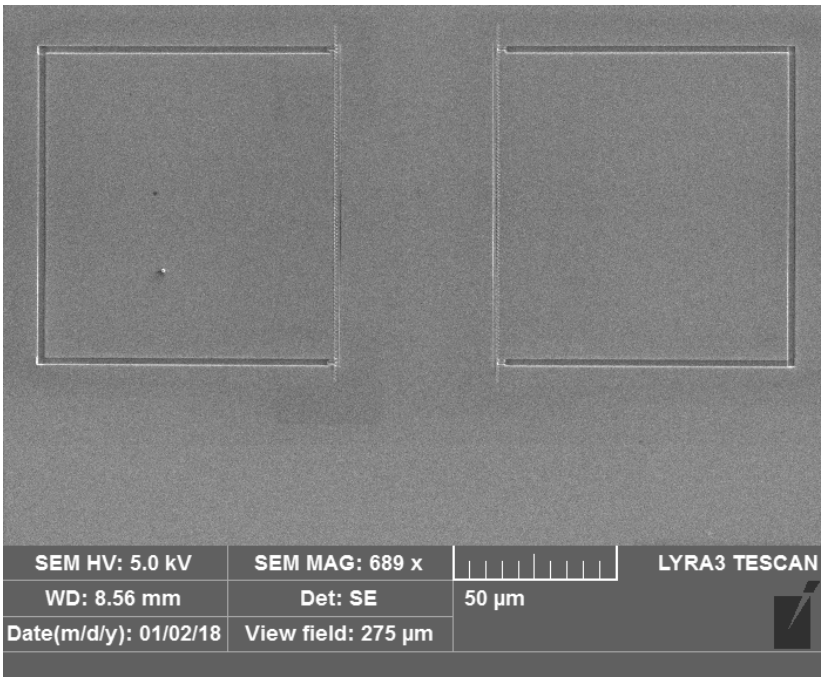


Figure 5.6: Pair of proof of concept devices. The diodes are facing the same direction, but the contact pads are on opposite sides.

After this was accomplished, I proceeded with the chemical APM etch referred to earlier. I left the sample in the chemical bath for two minutes to achieve the desired undercut, and then rinsed with de-ionized water. I then removed the hard mask with the ion gun shown in figure 5.2. I only required ten seconds of exposure to remove what was left of the hard mask, as the etch rate for aluminum is much higher than that of TiN [47]. I then took the processed wafer and conducted an AFM study to ensure the critical feature sizes and profile of the diodes were appropriate for the desired parameters. The AFM profile can be seen in figure 5.7. While the feature sizes are on the desired size scale, whether they are small enough to ensure proper rectification can only be determined by testing. The TiN film post processing is thin, on the order of 50nm. At 1.5 microns, this thickness has a high degree of transparency. As such, the effective coupling of the sample is substantially reduced. Simulations of the actual fabricated device indicate that the maximum power that can be obtained from the actually fabricated device is 0.2% with essentially no frequency discrimination.

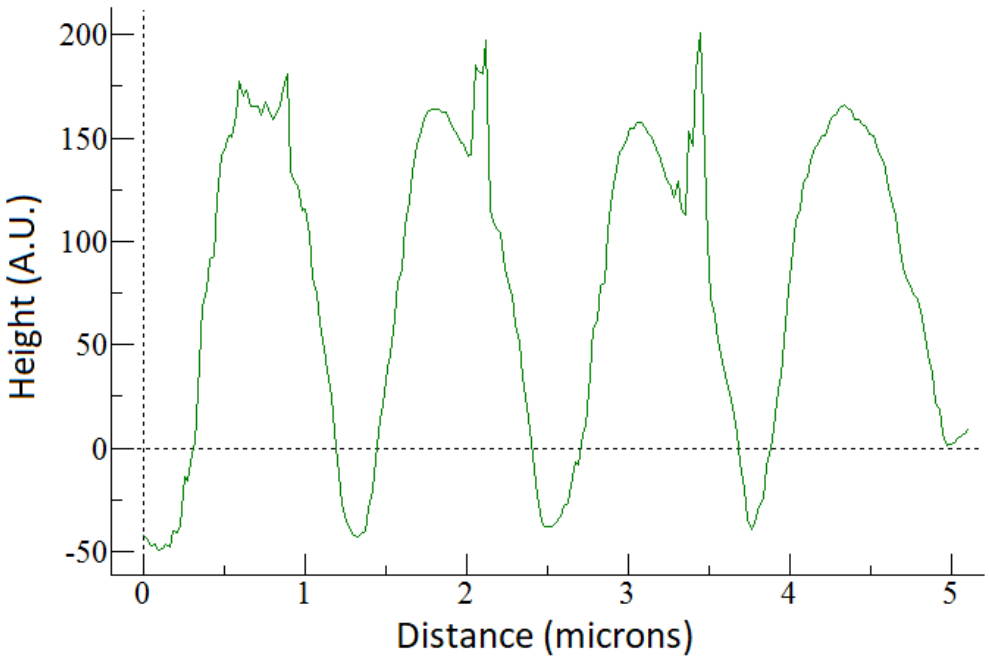
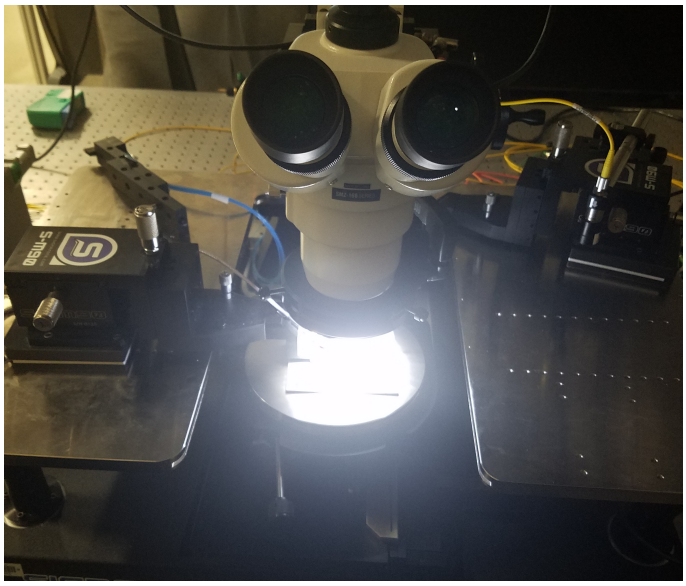


Figure 5.7: AFM scan of the critical feature size for the geometric diode array.

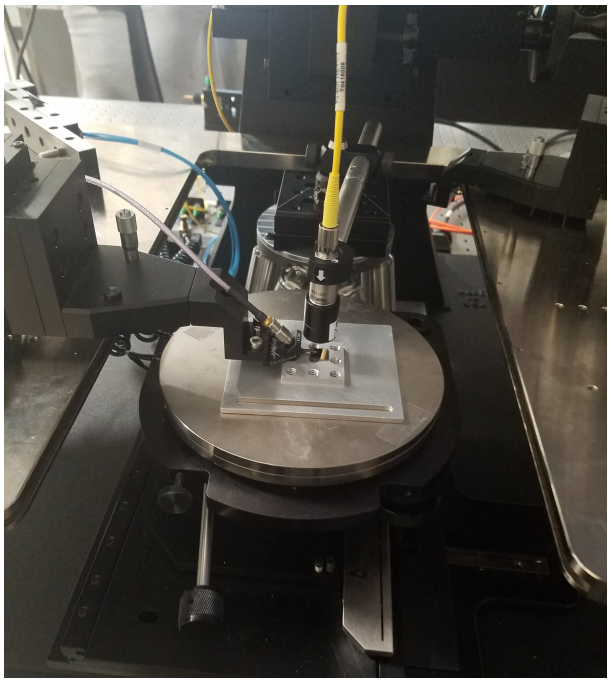
CHAPTER 6: RESULTS AND CONCLUSIONS

6.1 Testing Results

I expect to see two main effects from this device. First, I expect a DC offset due to a mixture of hot electrons generated by absorption events in the surface plasmon wave and from phase randomized mixing in the rectified signal. I also expect to see the input beat frequency resolved from the geometric diodes. I arranged the experiment to be able to, with various instruments, determine both of these quantities. My setup can be seen in figure 6.1. I utilized a master slave fiber coupled diode laser centered at 1.5 microns to generate the beat frequency, with a power output of 10mW. The beat was tunable from 500 MHz to 100GHz and verified via an optical spectrum analyzer. The fiber carrying the laser was coupled to a collimator, and a lens was placed approximately 12mm in front of the collimator to focus the beam on the surface. This was connected to an hexapod stage, capable of finely controlling position in x, y, z, and under rotation. To detect the signal, an RF probe was positioned on the contact pad, as seen in figure 6.2.As can be seen, the probe is approximately the same size as the contact pad, which resulted in capacitive effects and extreme sensitivity to position. In order to ensure the probe maintained proper contact with the pad while the microscope was exchanged for the fiber, I utilized a vector network analyzer to ensure the impedance maintained a nonzero profile (indicating there was not a short). Once the positioning was completed, the RF probe could then be hooked up to a piccoammeter to measure a DC current obtained via laser illumination. Whilst positioning was difficult, I was able to maintain a stable connection and make several measurements. One set of measurements under 10mW illumination provided a current of 530nA $\pm$80nA, with a noise floor of 120nA $\pm$50nA. Another set provided 680nA



(a) Microscope inserted setup



(b) Fiber inserted setup

Figure 6.1: (a) Setup for alignment of the RF probe on the sample via dark field microscope. Because of the microscope size and working distance, the fiber had to be removed and can be seen sitting to the right. (b) Setup with the fiber inserted via a magnetic holder on the hexapod stage. The light was focused onto the area near the RF probe and positioned via the hexapod stage to the optimum location.

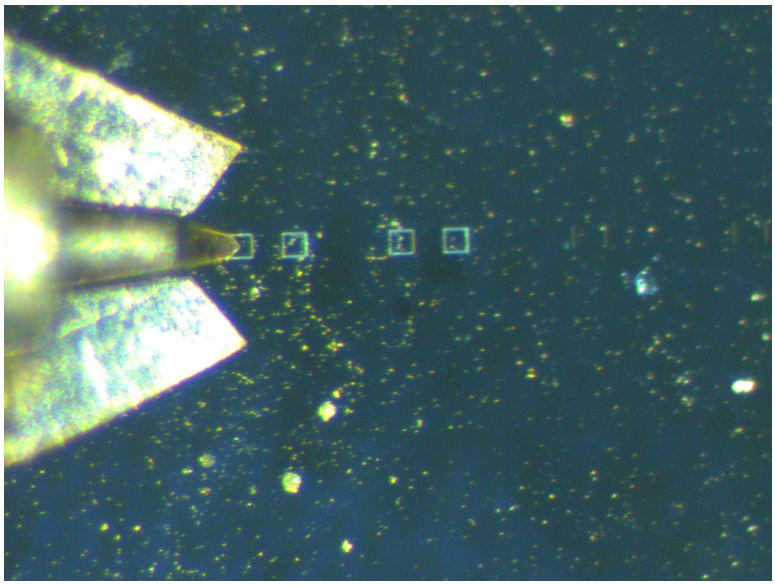


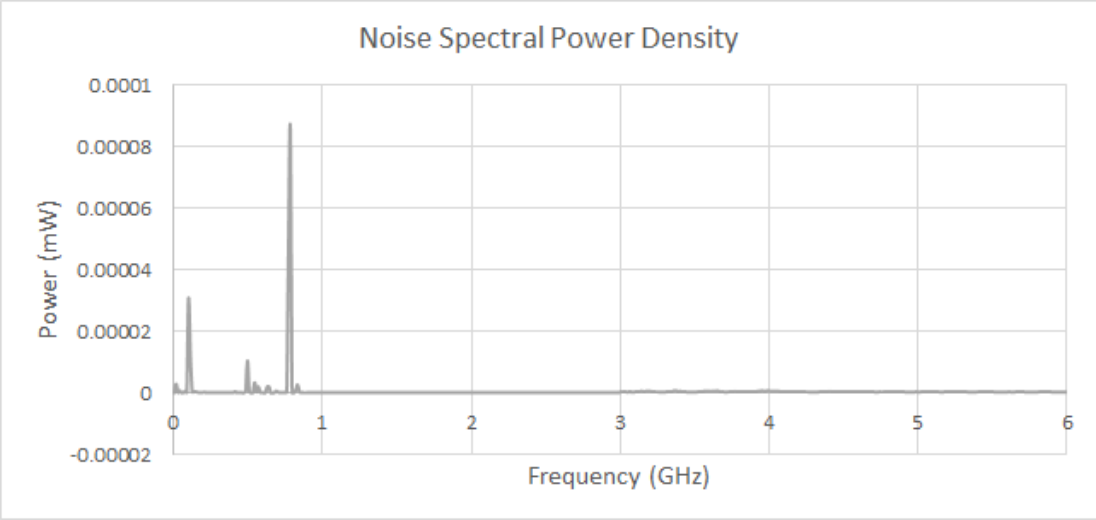
Figure 6.2: Image of the RF probe in contact with a contact pad of the fabricated device.

±80nA with a noise floor of 5nA ±5nA. The existence of the DC current would allow for passive power harvesting in the event the device were not being actively used. Although this is a small quantity, coupled with passive operation it would still be a net gain in the power usage of the detector.

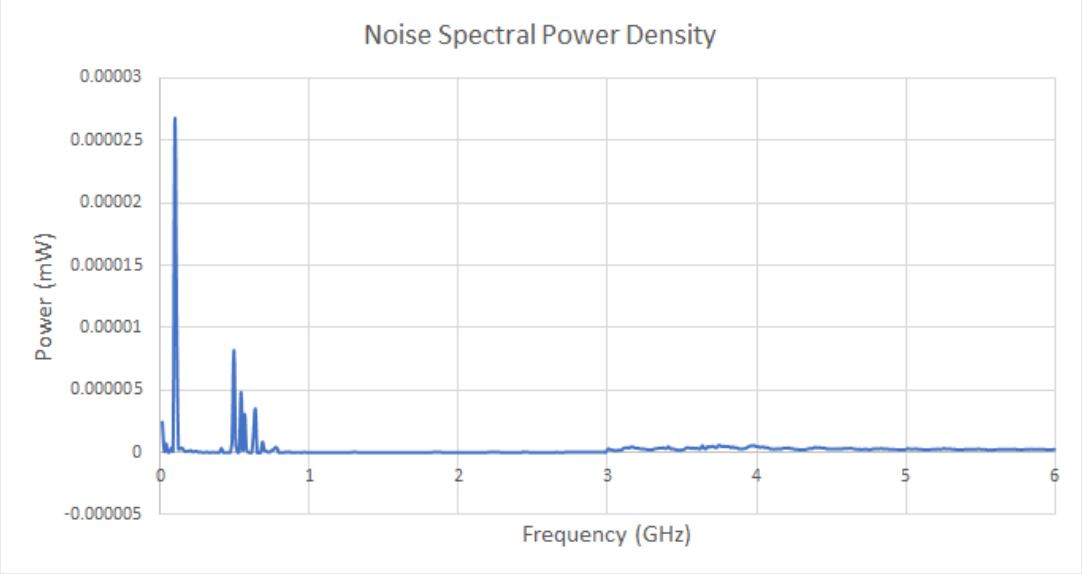
In addition to these measurements being well above the noise floor, the current disappeared when the laser was turned off or the fiber disconnected. I also attempted to obtain these results with the set of arrays where the diode direction was opposite, and I was unable to detect any statistically significant increase in DC current. I was able to, with great certainty, predict whether the laser was turned off or the fiber was disconnected solely from viewing the current response, thus qualitatively establishing the effectiveness of this device as a coherent source detector solely based upon the DC response.

In addition to the measurements at DC, I tested the frequency response by attaching the RF probe to a spectrum analyzer. In order to establish a baseline, I first compared the noise profile with the RF probe when the probe was not on

the contact pad with the profile of the probe on the contact pad. As expected, the contact pad was effectively an electrically small patch antenna. Since my setup was not shielded, this resulted in picking up several bands of broadcasted signal, such as television channels, cellular signals, and bluetooth (figure 6.3).



(a) Noise for 1.85GHz Beat Experiment



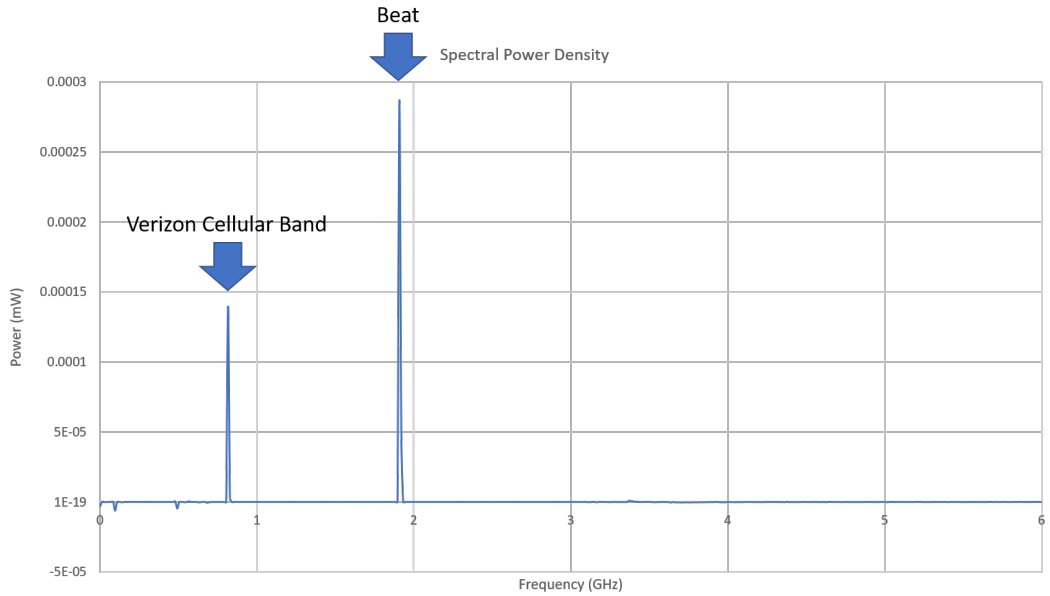
(b) Noise for 0.82GHz Beat Experiment

Figure 6.3: (a) Average of noise profiles taken for an experiment which would be illuminated with a 1.85GHz beat. (b) Average of noise profiles taken for an experiment which would be illuminated with a 0.82GHz beat.

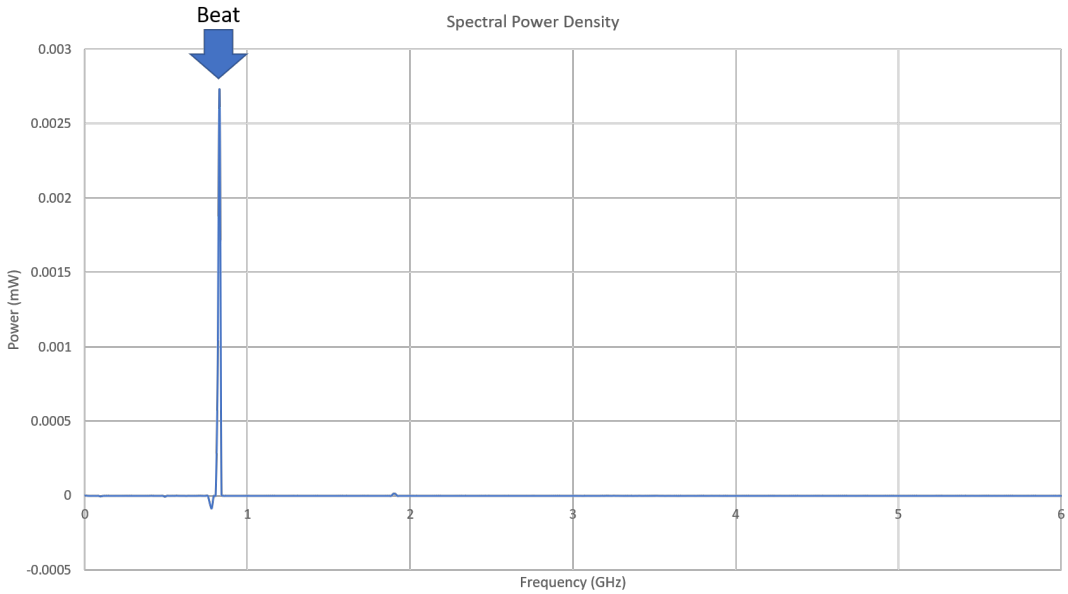
However, this aided us in ensuring proper contact of the probe with the pad as there was discrimination of position based on the background signal. In order to account for these spurious signals, several noise spectra were taken before each measurement and averaged to determine the background noise levels. To ensure a measurable signal was coupling to the RF probe, I needed to operate at a frequency which was impedance matched to the spectrum analyzer. Since there were strong capacitive effects due to the relative size of the RF probe to the contact pad, there were a limited number of frequencies available which had an impedance of 50Ω. The two frequencies we chose to operate at were 0.82GHz and 1.85GHZ. 0.82 GHz is used is primarily used in mobile telephone networks and, as can be seen in figure 6.4 (a), were detectable by the contact pads when a call was in progress in the nearby vicinity of our test. The strength of the recieved signal did not rise substantially above 150nW over the duration of the call however, and for further testing I ensured that all cellular devices in the region were turned off. This resulted in the band no longer appearing in our noise measurments (see figure 6.3 b). The 1.85GHz region is primarly used for earth to space communication and control, and as such I never observed any substantial peak in the noise values at this frequency. Through several measurements I was able to obtain signal at the beat frequency (verified by comparison to the optical spectrum analyzer for the laser). These results indicate that approximately 275nW are visible in the rectified frequency bands. Since the spectrum analyzer I was utilizing was not designed to detect DC, I could not simultaneously measure the frequency response and the DC current response, so I am unable to determine exactly what percentage of the incident power coupled to the slit. However, if I assume that the DC response and the beat frequency response both exist for the measurement, I can hypothesize that for a diode responsivity of 0.1A/W, approximately 8 μW of power was coupled to plasmons. If I assume that the 10 mW illumination only generated plasmons at the slit and was focused entirely on the slit, the simulations would indicate that 20μW

is the maximum detectable power from the device, indicating that this particular fabrication performed at a hypothetical 40% of maximum efficiency. Since the focal spot was not small enough or properly shaped to perfectly couple to only the slit, it is entirely likely that the possible power coupling was lower. In addition, plasmons could have been generated at the diode edges themselves, increasing the effective DC baseline. Thus I cannot say with certainty exactly what the proper efficiency of the device is, only that it falls within expected maximum parameters and produces a measurable response.

I have shown that I can measure the beat frequency, and the implications of this are significant. Although with this specific geometry I cannot use this beat to determine what the input frequencies were, this does indicate that it is possible to use this approach to obtain frequency information from the detector. This could be accomplished either via a frequency selective surface to launch the plasmons, or possibly via a reference beam to establish the beat, so long as the bandwidth is sufficiently narrow. I did not utilize an oscilloscope to determine the relative phase profile, but the existence of the narrowband frequency on the spectrum analyzer strongly implies the ability to obtain the relative phase information. As such, pursuing other geometries and coupling methodologies designed around obtaining this information, as well as characterizing power dependence and diode choice, are reasonable follow up experiments. These results, while by no means a final commercial device, indicate the feasibility of this approach as a detector which provides intensity and frequency information. This work is the first example of a combined on chip plasmonic heterodyne detector. The design and implementation of the detector, built upon a solid theoretical background and existing research base, promises to provide an a platform for obtaining more information from an electromagnetic field than existing detectors.



(a) 1.85GHz Beat



(b) 0.82GHz Beat

Figure 6.4: (a) Detector response for a laser beat frequency of 1.85GHz normalized to several noise measurements. The band at 0.82GHz is due to a nearby cell phone call on the verizon network taken after the noise readings. (b) Detector response for a laser beat frequency of 0.82GHz normalized to several noise measurements. I ensured no cell phones in the nearby area were turned on for any measurements.

6.2 Conclusions

The net result of this body of work encompasses design and fabrication of metasurfaces for field control and detection. I established a working methodology for designing metasurfaces based on a multistep simulation process in COMSOL. This approach is straightforward and typically required as a basis for most metasurface design applications that cannot be solved analytically. I also introduced the concept of utilizing voids or occlusions in specific patterns to generate a metasurface, enabling a new class of embedded metasurfaces. I demonstrated through rigorous full wave simulation that this design methodology can create functioning optical elements with reasonable efficiency. I then extended this approach to create a database backed program for designing holographic metasurfaces. This revolutionary program extends basic design principles of traditional holography to a discrete metasurface based implementation. This program takes any traditional image format as input and, through a series of reductions and transforms, maps a metasurface distribution that would produce the image on propagation. The program then encodes the physical dimensions of the metasurface based upon information in the database to a .dxf schematic which can be used in fabrication of the surface. This approach produces good quality results in substantially less time than would be required for metasurface hologram design via traditional methodologies once the database is set up. The database, although specific to material and wavelength, can be modified at will through the addition of resonator responses in a .csv file which can be directly uploaded via the graphic user interface. Thus, the program itself acts as a starting point from which to expand the design capabilities into whatever realm any enterprising company may wish to develop. This powerful tool enables a fast throughput metasurface design process that can improve the design and fabrication times for improved industrial use.

The development of methodologies and processes for design of metasurfaces naturally led into designing a metasurface for extended functionality such as field de-

tection. This development, having evolved through several stages of conceptual implementation, led to a comprehensive design process which can produce designs for detectors at any desired wavelength. These detectors have the capacity to extract relative phase information from the incident field, and in certain configurations can distinguish between coherent and incoherent radiation. They offer the advantage of operating fundamentally on a principle of heterodyning, which as a method of direct field detection eliminates the need for cooling in infrared detectors, except perhaps as a method to reduce thermal noise. By ensuring that the manufacturing process and materials are compatible with industrial methodology, this detector promises to be a useful addition to the field. Each of these advantages opens up a realm of application that are currently challenges for traditional photodiodes. First, the ability to relatively easily identify materials for operation at any wavelength allows for cheaper and less toxic detectors than are currently utilized. This will improve availability of detection in mid wave and long wave infrared. Hand in hand with this is the fact that the device can detect radiation at room temperatures which will reduce the need for cooling. This will reduce the bulk and cost of incorporating mid and long wave detectors into hardware by eliminating the coolant storage and circulation requirements. In addition, since the device doesn't strictly require a bias to operation, when not in use the rectified signal can be used as power harvesting. While this would be at low efficiency, it would improve power consumption and may reduce the battery load required for the operation of some devices and systems. All of those advantages are merely a solution to current difficulties with photodiode detectors. Where this approach really shines is in the relative phase detection capabilities inherent in the heterodyne based detection methodology. Since traditional photodiodes only detect energy from electron transitions, all phase information is lost in the detection process. This detector, while it can't determine absolute phase unless a reference source is used, is capable of detecting the relative phase between two adjacent fre-

quencies. Although this capability exists compactly at low frequency and in bulky setups at high frequency, this is the first feasible on chip solution that integrates detection and phase information for high frequencies. This makes it possible to utilize complex signal analysis of high frequency radiation for device level implementation. Applications ranging from astronomy to wavefront analysis abound, possible through implementation of this device.

6.3 Future Work

As with all of science, there are always more questions to ask and more problems to solve, and this work is no exception. Designing metasurfaces is unique to the team designing them. What assembly of analytics, simulation, and algorithm is used depends on the research team doing the designing. My approach is predominantly simulation and algorithm, and so the work done to enhance the program will be focused in these areas. The most obvious enhancement to the program is to expand the database to include resonator profiles for different wavelengths and materials. This is merely a matter of time and simulation, and is a straightforward way to improve functionality of the program. The second way to improve the program is to implement an iterative Fourier transform approach such as the Gerchberg-Saxton algorithm from [48] or similar works. The goal with this would be two fold. First, it should reconstruct the phase that produced the image in the first place, making the generated hologram more realistic. Second, it should remove the noise from the holographic image. Whilst my reduction approach helped to reduce this noise, it can be further reduced to provide a higher quality hologram. This would, of course, be a trade off with time as the iterative loop would increase computation time. However, since the current time sink in the program is mapping the resonators to the matrix, this may be an acceptable increase in computation time. Further improvements could be made on enhancing the user interface and enabling the export of different file formats as well, in order to improve user friendliness. In addition, by utilizing the

same form of program, it is possible to assemble a database of detector elements, which could then be converted into a detector plane by a similar methodology to the hologram generation. This would allow for reasonably efficient and quick focal plane array design.

The detector, being a new device, still has a good bit of work to do to bring it to a high enough technology level to be commercializable. Likely the biggest benefit would come from further exploration into diode integration and structuring. At near and mid IR, the materials are fairly straightforward for diode integration, but at long wave IR, the conductivity of the semiconductors varies only slightly from that of unintentionally doped semiconductors (from an electrons in the conduction band perspective. Thus, forming a schottky or metal insulator metal diode is impractical. In these cases, a PN junction would have to be used to rectify the current of the plasmon. This, and the particular materials to use for diodes at various frequencies such that there is proper integration into the detector is a large enough project for another dissertation, but would be the next logical step to raise the technology level up to the pre-prototype stage. Once that has been accomplished, then work can begin in earnest to develop a readout schematic for a full array of these diodes. Establishing a proper focal plane array design would move the research into the prototype stage. Each individual pixel would need to output its rectified signal, which would need to be mixed with other pixels of the same sensitivity whilst maintaining any relative phase information in the signal. These would need to then be read out and information extracted. This layout would allow engineering optimizations and programming of the electronics reading the signal to bring the work to a commercial level. A full focal plane with readout, ready for introduction into cameras and scientific instruments all over the world. Each of these steps is a major research task, but would push the door open to a fully functional device ready to fill a need and open new and exciting areas in detectors and direct field detection.

REFERENCES

[1] A. I. Fernández-Domínguez, F. J. García-Vidal, and L. Martín-Moreno, "Unrelenting plasmons," *Nature Photonics*, vol. 11, pp. 8–10, 2017.

[2] L. Marrucci, C. Manzo, and D. Paparo, "Pancharatnam-berry phase optical elements for wave front shaping in the visible domain: switchable helical mode generation," *Applied Physics Letters*, vol. 88, no. 22, p. 221102, 2006.

[3] D. Lin, P. Fan, E. Hasman, and M. L. Brongersma, "Dielectric gradient metasurface optical elements," *science*, vol. 345, no. 6194, pp. 298–302, 2014.

[4] N. Yu, P. Genevet, M. A. Kats, F. Aieta, J.-P. Tetienne, F. Capasso, and Z. Gaburro, "Light propagation with phase discontinuities: generalized laws of reflection and refraction," *science*, vol. 334, no. 6054, pp. 333–337, 2011.

[5] P. Genevet, N. Yu, F. Aieta, J. Lin, M. A. Kats, R. Blanchard, M. O. Scully, Z. Gaburro, and F. Capasso, "Ultra-thin plasmonic optical vortex plate based on phase discontinuities," *Applied Physics Letters*, vol. 100, no. 1, p. 013101, 2012.

[6] S. B. Glybovski, S. A. Tretyakov, P. A. Belov, Y. S. Kivshar, and C. R. Simovski, "Metasurfaces: From microwaves to visible," *Physics Reports*, vol. 634, pp. 1–72, 2016.

[7] X. Luo, M. Pu, X. Ma, and X. Li, "Taming the electromagnetic boundaries via metasurfaces: from theory and fabrication to functional devices," *International Journal of Antennas and Propagation*, vol. 2015, 2015.

[8] W. E. Kock, *Radar, sonar, and holography: an introduction*. Academic press, 2013.

[9] B. A. Munk, *Frequency selective surfaces: theory and design*, vol. 29. Wiley Online Library, 2000.

[10] M. Fox, *Optical properties of solids*, vol. 3. Oxford university press, 2010.

[11] A. D. Taylor, J. Ladd, Q. Yu, S. Chen, J. Homola, and S. Jiang, "Quantitative and simultaneous detection of four foodborne bacterial pathogens with a multichannel spr sensor," *Biosensors and Bioelectronics*, vol. 22, no. 5, pp. 752–758, 2006.

[12] H. Hu, D. Ji, X. Zeng, K. Liu, and Q. Gan, "Rainbow trapping in hyperbolic metamaterial waveguide," in *CLEO: QELS_Fundamental Science*, pp. QTu2A–4, Optical Society of America, 2013.

[13] Z. Naqvi, C. Rosenbury, M. A. Fiddy, and T.-H. Her, "Fiber-metasurface for wavefront shaping," in *CLEO: Applications and Technology*, pp. JTu5A–48, Optical Society of America, 2017.

[14] W. T. Chen, A. Y. Zhu, V. Sanjeev, M. Khorasaninejad, Z. Shi, E. Lee, and F. Capasso, "A broadband achromatic metalens for focusing and imaging in the visible," *Nature nanotechnology*, p. 1, 2018.

[15] M. D. Huntington, L. J. Lauhon, and T. W. Odom, "Subwavelength lattice optics by evolutionary design," *Nano letters*, vol. 14, no. 12, pp. 7195–7200, 2014.

[16] S. A. Ponomarenko and E. Wolf, "Spectral anomalies in a fraunhofer diffraction pattern," *Optics letters*, vol. 27, no. 14, pp. 1211–1213, 2002.

[17] L. Song, R. A. Lessard, and P. Galarneau, "Diffraction efficiency of a thin amplitude-phase holographic grating: a convolution approach," *Journal of Modern Optics*, vol. 37, no. 8, pp. 1319–1328, 1990.

[18] R. W. Gerchberg, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[19] F. J. Salgado-Remacha, "Reducing the variability in random-phase initialized gerchberg-saxton algorithm," *Optics & Laser Technology*, vol. 85, pp. 30–34, 2016.

[20] H. Raether, "Surface plasmons on smooth surfaces," in *Surface plasmons on smooth and rough surfaces and on gratings*, pp. 4–39, Springer, 1988.

[21] S. Nelson, K. S. Johnston, and S. S. Yee, "High sensitivity surface plasmon resonace sensor based on phase detection," *Sensors and actuators B: Chemical*, vol. 35, no. 1-3, pp. 187–191, 1996.

[22] T. Aihara, K. Nakagawa, M. Fukuhara, Y. Ling Yu, K. Yamaguchi, and M. Fukuda, "Optical frequency signal detection through surface plasmon polaritons," *Applied Physics Letters*, vol. 99, no. 4, p. 043111, 2011.

[23] C.-W. Chang, C.-E. Lin, C.-J. Yu, T.-T. Yeh, and T.-J. Yen, "Miniature surface plasmon polariton amplitude modulator by beat frequency and polarization control," *Scientific reports*, vol. 6, 2016.

[24] D. Choi, C. K. Shin, D. Yoon, D. S. Chung, Y. W. Jin, and L. P. Lee, "Plasmonic optical interference," *Nano letters*, vol. 14, no. 6, pp. 3374–3381, 2014.

[25] J. S. Fakonas, H. Lee, Y. A. Kelaita, and H. A. Atwater, "Two-plasmon quantum interference," *Nature Photonics*, vol. 8, no. 4, pp. 317–320, 2014.

[26] X. Yang, B. Zeng, C. Wang, and X. Luo, "Breaking the feature sizes down to sub-22 nm by plasmonic interference lithography using dielectric-metal multilayer," *Optics express*, vol. 17, no. 24, pp. 21560–21565, 2009.

[27] N. Liu, L. Langguth, T. Weiss, J. Kästel, M. Fleischhauer, T. Pfau, and H. Giessen, "Plasmonic analogue of electromagnetically induced transparency at the drude damping limit," *Nature materials*, vol. 8, no. 9, pp. 758–762, 2009.

[28] Y. Wang, "Devices based on an array of light-filtering channels with surface plasmon interference filters," Jan. 31 2006. US Patent 6,992,826.

[29] X. Cui, X. Heng, C. Yang, A. Scherer, and D. Psaltis, "On-chip phase microscope/beam profiler based on differential interference contrast and/or surface plasmon assisted interference," Aug. 3 2010. US Patent 7,768,654.

[30] H. Wei, Z. Wang, X. Tian, M. Käll, and H. Xu, "Cascaded logic gates in nanophotonic plasmon networks," *Nature Communications*, vol. 2, p. 387, 2011.

[31] G. V. Naik, V. M. Shalaev, and A. Boltasseva, "Alternative plasmonic materials: beyond gold and silver," *Advanced Materials*, vol. 25, no. 24, pp. 3264–3294, 2013.

[32] U. Guler, A. V. Kildishev, A. Boltasseva, and V. M. Shalaev, "Plasmonics on the slope of enlightenment: the role of transition metal nitrides," *Faraday discussions*, vol. 178, pp. 71–86, 2015.

[33] Z. Zhu, S. Joshi, S. Grover, and G. Moddel, "Graphene geometric diodes for terahertz rectennas," *Journal of Physics D: Applied Physics*, vol. 46, no. 18, p. 185101, 2013.

[34] S. Grover and G. Moddel, "Applicability of metal/insulator/metal (mim) diodes to solar rectennas," *IEEE Journal of Photovoltaics*, vol. 1, no. 1, pp. 78–83, 2011.

[35] P. Periasamy, H. L. Guthrey, A. I. Abdulagatov, P. F. Ndione, J. J. Berry, D. S. Ginley, S. M. George, P. A. Parilla, and R. P. O'Hayre, "Metal–insulator–metal diodes: Role of the insulator layer on the rectification performance," *Advanced Materials*, vol. 25, no. 9, pp. 1301–1308, 2013.

[36] S. Cheung and N. Cheung, "Extraction of schottky diode parameters from forward current-voltage characteristics," *Applied Physics Letters*, vol. 49, no. 2, pp. 85–87, 1986.

[37] L. Liu, J. L. Hesler, H. Xu, A. W. Lichtenberger, and R. M. Weikle, "A broadband quasi-optical terahertz detector utilizing a zero bias schottky diode," *IEEE Microwave and Wireless Components Letters*, vol. 20, no. 9, pp. 504–506, 2010.

[38] R. Han, Y. Zhang, Y. Kim, D. Y. Kim, H. Shichijo, E. Afshari, *et al.*, "Active terahertz imaging using schottky diodes in cmos: Array and 860-ghz pixel," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 10, pp. 2296–2308, 2013.

[39] J. Hou, "Terahertz schottky diode device with high cutoff frequency developed in imecas," Nov 2012.

[40] H. Nejati and A. Beirami, "Theoretical analysis of the characteristic impedance in metal–insulator–metal plasmonic transmission lines," *Optics letters*, vol. 37, no. 6, pp. 1050–1052, 2012.

[41] K. D. Stock, "Internal quantum efficiency of ge photodiodes," *Applied optics*, vol. 27, no. 1, pp. 12–14, 1988.

[42] J. Hu, X. Ren, A. N. Reed, T. Reese, D. Rhee, B. Howe, L. J. Lauhon, A. M. Urbas, and T. W. Odom, "Evolutionary design and prototyping of single crystalline titanium nitride lattice optics," *ACS Photonics*, vol. 4, no. 3, pp. 606–612, 2017.

[43] M. Noginov, L. Gu, J. Livenere, G. Zhu, A. Pradhan, R. Mundle, M. Bahoura, Y. A. Barnakov, and V. Podolskiy, "Transparent conductive oxides: Plasmonic materials for telecom wavelengths," *Applied Physics Letters*, vol. 99, no. 2, p. 021101, 2011.

[44] F. Scotognella, G. Della Valle, A. R. S. Kandada, M. Zavelani-Rossi, S. Longhi, G. Lanzani, and F. Tassone, "Plasmonics in heavily-doped semiconductor nanocrystals," *The European Physical Journal B*, vol. 86, no. 4, p. 154, 2013.

[45] S. Law, L. Yu, A. Rosenberg, and D. Wasserman, "All-semiconductor plasmonic nanoantennas for infrared sensing," *Nano letters*, vol. 13, no. 9, pp. 4569–4574, 2013.

[46] M. D. Levenson, N. Viswanathan, and R. A. Simpson, "Improving resolution in photolithography with a phase-shifting mask," *IEEE Transactions on electron devices*, vol. 29, no. 12, pp. 1828–1836, 1982.

[47] K. R. Williams, K. Gupta, and M. Wasilik, "Etch rates for micromachining processing-part ii," *Journal of microelectromechanical systems*, vol. 12, no. 6, pp. 761–778, 2003.

[48] F. Wyrowski and O. Bryngdahl, "Iterative fourier-transform algorithm applied to computer holography," *JOSA A*, vol. 5, no. 7, pp. 1058–1065, 1988.