NANO C-APERTURES APPLIED TO NEAR-FIELD INSPECTION AND DATA STORAGE

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF APPLIED PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

John Brian Leen
March 2010
© 2010 by John Brian Leen. All Rights Reserved.
Re-distributed by Stanford University under license with the author.

This work is licensed under a Creative Commons Attribution-
Noncommercial 3.0 United States License.
http://creativecommons.org/licenses/by-nc/3.0/us/

This dissertation is online at: http://purl.stanford.edu/zq719rc2501
I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Lambertus Hesselink, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

David Miller

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

R Pease

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.
Abstract

The restriction of optical spots sizes to the diffraction limit of \( \sim \lambda/2 \) is a long standing problem in optical imaging and optical data storage. As the length scales of interest to scientists and engineers have decreased, the diffraction limit has become a very significant issue and the need for a way to efficiently focus light to spot sizes of \(<\lambda/10\) has become more pressing. Near-field optical techniques are the only solution to optical probing at the deeply subwavelength scale, but the existing methods of solid immersion lenses (SILs), sharp probe tips (ANSOM) and subwavelength apertures (NSOM) have all been hamstrung by significant deficiencies. Specifically, SILs are limited by the available materials with high refractive indices, ANSOM suffers from high background signal and thus poor SNR and apertured NSOM is plagued by extremely low aperture transmission efficiencies that fall off as \((r/\lambda)^4\).

In 2002 Xiaolei Shi, a previous member of our research group, proposed the C-aperture as a superior near-field optical source and FDTD simulations showed that the aperture was a resonant structure capable of producing very small and very intense near-field optical spots. Many additional simulation studies followed and experimental work measuring far-field transmission verified that the aperture did, in fact, exhibit a wavelength resonance and near-field scanning optical microscopy studies showed that it confined light to a small area in the near-field.
I have continued this experimental work by first demonstrating that existing methods of fabricating C-apertures are incapable of producing C-apertures that fully realize the high, absolute field intensities promised by simulations. Using a new silicon nitride membrane based fabrication method, I showed that the performance of the C-aperture can be greatly enhanced both in the near and far-field by more efficiently concentrating light at the end of the ridge tip. This results in a near-field intensity more than an order of magnitude greater than the incident light and provides an improvement over conventional C-aperture performance of 8.8x higher intensity in the far-field and 630x higher in the near-field. I showed that this small and intense near-field spot can be used for near-field optical imaging with a subwavelength resolution of \( \approx \lambda/10 \). Finally, I demonstrated that the C-aperture manufactured on a silicon nitride membrane can be used to record and read data bits in a phase change recording medium (Ge\(_2\)Sb\(_2\)Te\(_3\)) with bit sizes of 53.5x50.2 nm or \( \approx \lambda/20 \).
Acknowledgements

There are many people who have been instrumental to the progress and completion of this work and whom I would like to thank for their help. First, and most obviously, is my advisor Bert Hesselink to whom I’m very grateful. None of this would have been possible without his advice and his efforts to procure the necessary funds to support my research. Although he provided excellent input throughout my work, he has a stated, intentional and deliberate hands-off style of leadership that can at times be frustrating, but in the end, I believe that it has made me much more self sufficient, capable of running complicated research without oversight. I am very glad to have this skill.

I’d also like to thank the other members of the Hesselink lab for their conversation, help and guidance throughout my years at Stanford. Notably Xioalei Shi and Joe Matteo, both past students who were very helpful in getting my work started and getting me up to date on the state-of-the-art of C-apertures. I would especially like to acknowledge my co-authors, Paul Hansen and Yao-te Cheng. Paul is the single soul responsible for creating the FDTD code ‘Trogdor’; without it I would not have been able to perform any of the realistic simulations that were crucial to designing and understanding the experiments in this dissertation. Without Yao-te, I believe that I would have spent many more years developing usable C-apertures than I already have. Yao-te is tenacious. Not only did he fabricate the membranes and chrome nano-wires used in this work but he pushed me to resurrect and
examine the use of membranes for aperture manufacturing, an idea that I had discarded long ago as too fraught with complications.

Outside the Hesselink lab, I owe a special thanks to Robert Barretto of the Schnitzer lab for his help in access to, and use of, their Ti:saph laser that was instrumental to the TPA work. I’d also like to thank Aaron Gibby, who was extremely generous with his help in deposition of the Ge$_2$Sb$_2$Te$_5$ recording media.

I’d like to thank the SNF staff for their help throughout, especially Mahnaz Mansourpour who was extremely helpful in advising on and accommodating the very non-standard fabrication processes I tried and developed. Another facility critical to my work was the physics machine shop and I’d like to thank the guys there for their help, especially Mehmet Solyali who is a gifted teacher of the machining arts. Also, of all the people who had no obligation to go out of their way to help me but routinely did so anyway, Paula Perron of the Applied Physics Department office is a standout.

My good friends Xiaobo Yin and Dan Witte were indispensable, both for their camaraderie and for the excellent advice they provided on all matters of experimental research. Dan Pickard falls into the same category of friend and sage, with the added point that he generously provided a great deal of critical equipment at critical times (with the consent of Fabian Pease, for which I am grateful). To XB, Dan and Dan, I am indebted.

I would also like to acknowledge the various institutions that provided the funding for this work, specifically the National Science Foundation, the General Electric Corporation, the Samsung Corporation and the KLA-Tencor Corporation.

Finally, I am very grateful to my family for their support all these years. I am grateful to my mom and dad who have given me so much and put me in a position to pursue and successfully complete this work. I am thankful for the simple presence of my son, Sam, who has brought a heaping portion of delight to the last 23 months. Lastly, I thank my wife,
Emilee, for her support. She is a consummate cheerleader and I am lucky to have had her help and love before and during my time in the Applied Physics Ph.D. program. It is unreasonable to say that without her, I could not have done this, but I can say that it would have been much less pleasant. Together, Emilee and Sam have made the times of defeat more tolerable and the times of success all the sweeter.

J. Brian Leen

Stanford University

March, 2010
Table of Contents

Abstract ......................................................................................................................... iv
Acknowledgements ....................................................................................................... vi
List of Tables ............................................................................................................... xii
List of Figures .............................................................................................................. xii
1 Introduction ........................................................................................................ 1
  1.1 The Optical Near-Field and the NSOM ........................................................... 2
  1.2 Limitations of Round Apertures ...................................................................... 6
  1.3 The Apertureless NSOM ................................................................................. 8
  1.4 Enhanced Transmission Apertures ................................................................. 10
    1.4.1 Aperture Arrays ........................................................................................ 10
    1.4.2 Transmission Mode Apertures ................................................................. 13
      A. Aperture and Antenna Fabrication ............................................................ 14
      B. Bowties ..................................................................................................... 16
      C. C and H-Apertures .................................................................................... 18
        Mode Structure and Optical Confinement .................................................... 19
        Waveguide vs. Aperture ............................................................................. 21
        Resonant Transmission ............................................................................. 24
        Input and Output Coupling ...................................................................... 26
        Summary of Important C-Aperture Properties ......................................... 29
  1.5 My Contributions and Thesis Outline ............................................................ 29
2 Aperture Fabrication Using Silicon Nitride Membranes ........................................ 32
  2.1 Introduction .................................................................................................... 32
  2.2 Sidewall Taper and Edge Rounding .............................................................. 33
  2.3 Gallium Contamination ................................................................................. 34
  2.4 Through Membrane Milling (TMM) ............................................................... 34
  2.5 Sample Fabrication ........................................................................................ 37
  2.6 Simulation ...................................................................................................... 38
  2.7 Experimental Verification ............................................................................. 41
  2.8 Summary ........................................................................................................ 44
3 TPA Probe of the C-Aperture Near-Field .............................................................. 46
Appendix ..........................................................................................................................92

7.1 Finite Difference Time Domain (FDTD) Simulations ...........................................92
  7.1.1 Background and Techniques ....................................................................92
  7.1.2 The Simulation Space .............................................................................94
  7.1.3 Analysis ....................................................................................................96
    A. Simulation Output .....................................................................................96
    B. Electric Field Distributions .....................................................................97
    C. Far-Field Transmission ...........................................................................98

7.2 C-NSOM Operation ............................................................................................101
  7.2.1 Mechanical Construction .......................................................................101
  7.2.2 Tip Assembly .........................................................................................103
  7.2.3 Scanning Methodology ..........................................................................104

Bibliography ................................................................................................................107
List of Tables

Table 4.1 – Measured and simulated transmission dip FWHM for various nano-wire widths in the TM and TE orientation. *For the 50 nm wire in the TE orientation, width is the distance between the right hand half minimum of the dip and the left hand half maximum of the peak ................................. 65

List of Figures

Figure 1.1 – (a) Conceptual NSOM design operating in illumination-transmission mode. (b) Operating in collection mode. (c) Operating in illumination-reflection mode........ 4

Figure 1.2 – (a) An early NSOM probe fabricated by heating and pulling an optical fiber. (b) Aperture at the end of the fiber produced by rotation of a pulled fiber during Al vapor deposition. (c) Configuration used for coating optical fibers with Al to produce a small, round aperture at the tip. Images reproduced from [12]. ............... 5

Figure 1.3 – A plot of power throughput vs. aperture radius for $\lambda = 1\mu$m. Dots are finite difference time domain simulations of round apertures in an infinitesimally thin perfectly conducting screen. Solid line is the Bethe power law. Plot reproduced from reference [14]. ................................................................. 6

Figure 1.4 – FDTD simulation of a 50 nm round aperture near an optical data storage medium. Side lobes are due to leakage of the illumination light at the edges of the Drude metal and perfectly matched layer (PML). .................................................. 8

Figure 1.5 – Apertureless NSOM configuration................................................................. 9

Figure 1.6 – E-beam vs. FIB nano-structure fabrication. (a) e-beam lithography of a resist. (b) Metal deposition. (c) Lift-off of the unwanted metal. (d) Physical etching using a gallium ion beam. (e) The resulting pit with ion damage and material mixing. ........................................................................................................... 15

Figure 1.7 – Bowtie configurations: (a) Bowtie antenna. Polarization of the illuminating field is vertical. (b) Bowtie aperture. Polarization of the illuminating field is horizontal. The gap between bowtie tips is g and the radius of curvature of the
tips is r. +/- signs represent displaced charge in the metal during operation. Dotted arrows indicate the direction of surface currents.

Figure 1.8 – (a) Diagram of the C-aperture (single ridge waveguide). (b) Diagram of the H (or I) aperture (double ridge waveguide).

Figure 1.9 – Plot of the waveguide cutoff wavelength normalized to a (see Figure 1.8). The variables a, s, d and b refer to the dimensions in Figure 1.8. Image reproduced from Ref [43].

Figure 1.10 – Electric field distributions in infinite PEC waveguides at $\lambda = 980$ nm (scale bar under (a)). The colormap shows electric field intensity and arrows trace electric field lines, showing the proportional electric x-y field. (a) The standard C1 aperture design from Ref [14]. (b) Same as (a) but with the ridge tips pulled together to form a tip. (c) Same as (a) but with the ridge corners rounded with a radius of curvature 10 nm. (d) The H-aperture from Ref [14]. (e) Same as (d) but with the ridge tips pulled together to form a tip. (f) Same as (e) but with the ridge base point removed to form a bowtie aperture.

Figure 1.11 – Maximum transmitted intensity for PEC waveguides of varying length. Curves entering from the left are Fabry-Perot like higher order resonant modes. Image reproduced from Ref. [48].

Figure 1.12 – Surface plasmon generation in nano-apertures. Grating matching occurs at the input and output edges making the corners the only source of SPs. (1) SPs propagating away from the aperture. (2) SPs propagating through the aperture. (3) SPs sourced at the aperture exit either by (2) or by a waveguide like mode and propagating away from the aperture.

Figure 1.13 – Example transmission resonance curve for a C-aperture in a 100 nm thick gold layer on a 75 nm thick silicon nitride membrane. Aperture size parameters: a = 195 nm, b = 145 nm, d = 50 nm, s = 65 nm.

Figure 1.14 – Far-field optical transmission through a variety of nano-apertures. Aperture shapes are shown in the left hand column. Incident light is white. (a) Transmission for horizontally polarized incident light. (b) Transmission for vertically polarized incident light. Image reproduced from [58].
Figure 1.15 – Near-field spot profiles produced by a C-aperture in a 100 nm thick gold layer on a 75 nm thick silicon nitride membrane. Aperture size parameters: a = 195 nm, b = 145 nm, d = 50 nm, s = 65 nm (identical to Figure 1.13). (a) Spot profile immediately adjacent to the metal surface. (b) Spot profile 18 nm from the metal surface. ................................................................. 27

Figure 1.16 – Spot evolution vs. propagation distance from the aperture exit for the same aperture as in Figure 1.13 and Figure 1.15. (a) x and y FWHM. (b) Spot maximum intensity. ................................................................. 28

Figure 2.1 – Diagrams of the 3 primary structural non-idealities caused by FIB milling. (a) Ga⁺ beam tails cause surface rounding and angularly dependent sputter yield causes side wall taper. (b) The Ga⁺ beam profile causes interior corner rounding.... 34

Figure 2.2 - Schematic drawing of (a) the top view of the mill pattern used in FIB milling and simulation, (b) a sample of the simulated gold surface roughness, (c) a cross section through the tongue of a DMM aperture with AuGa₂ lining, and (d) a cross section through the tongue of a TMM aperture. Optical illumination is from below and collection from above in (c) and (d) with measurement of the FDTD calculated electric and magnetic fields at the dotted plane........................................ 36

Figure 2.3 - SEM images viewed at 52° of tilt of (a) a DMM aperture and (b) a TMM aperture. AFM images of (c) a DMM aperture and (d) a TMM aperture. ............... 37

Figure 2.4 – (a) The simulated electric field intensity spectrum 2nm from the metal plane for a DMM aperture with AuGa₂ contamination. (b) Same as (a) but for a TMM aperture. (c) The simulated electric field intensity profile 18nm from the metal surface for a DMM aperture with AuGa₂ layer and resonant at 980 nm. The overlay in black (white) shows the aperture perimeter at the entrance (exit) of the metal layer. (d) Same as (c) but for a TMM aperture. The entrance and exit are nearly identical and shown in the black overlay. (e) x and y near-field spot FWHM values vs. propagation distance for DMM and TMM apertures. (f) Maximum electric field intensity vs. propagation distance for DMM and TMM apertures................................................................. 39

Figure 2.5 – Aperture far-field transmission scans for apertures ranging in size from α = 40-100 nm. (a) DMM horizontally polarized incident light. (b) TMM
horizontally polarized incident light. (c) DMM vertically polarized incident light. (d) TMM vertically polarized incident light. .................................41

Figure 2.6 - Far-field transmission for the two aperture types normalized to the peak value of the TMM curve. Experimental data are shown as solid curves and simulated data are shown as dashed curves: (a) TMM experimental data, (b) TMM simulated, (c) DMM experimental data, (d) DMM experimental data scaled by a factor of five, (e) DMM simulated without gallium contamination and (f) DMM simulated with gallium contamination. Horizontal error bars show the shift in simulated peak position for a 20% error in sputter yields. .................................42

Figure 3.1 – Diagram of the TPA near-field probe setup. .................................................................49

Figure 3.2 – (a) Aperture transmission spectrum in the presence of SU-8. (b) Maximum near-field intensity at the laser wavelength of 780 nm. (c) Intensity contour lines along the central x-z slice at the aperture exit. The aperture size is $\alpha = 45$ nm. (d) Intensity contour lines along the central y-z slice at the aperture exit. The aperture size is $\alpha = 45$ nm. ...........................................................................................50

Figure 3.3 – (a) AFM image of the polymerized spot produced by an $\alpha = 45$ nm C-aperture. $P_{inc} = 2.5$ mW. (b) AFM image of the polymerized spot from an $\alpha = 80$ nm C-aperture. $P_{inc} = 2.5$ mW. (c) AFM image of a 200 nm diameter circular aperture showing no polymerization. Area is ~3x the area of the $\alpha = 45$ nm C. $P_{inc} = 2.5$ mW. (d) AFM image of an $\alpha = 60$ nm DMM C-aperture with no polymerized spot at the ridge. $P_{inc} = 27$ mW. .....................................................................................................53

Figure 4.1 – (a) Diagram of the C-aperture dimensions. (b) SEM of a fabricated C-NSOM tip that has been cross sectioned by FIB to show the construction. The corresponding area of (b) is shown in the dashed black line. (c) Diagram of the C-NSOM. Illumination is through the optical fiber at the top. The silicon nitride membrane with truncated pyramidal tip and C-aperture is glued to the cleaved fiber end. The sample to be scanned is positioned under the C-NSOM tip and transmitted light is collected by a microscope objective. The red dotted box shows the volume included in FDTD simulations........................................................60

Figure 4.2 – Fabrication of a silicon nitride membrane with truncated pyramid. (a) A double polished <100> silicon wafer is LPCVD coated with 190 nm of low stress
silicon nitride. (b) A 5 µm square hole is lithographically patterned and etched in
the top surface using an O₂ and SF₆ RF plasma etch. (c) The wafer is immersed in
40% KOH in water at 70 °C to produce a pyramidal mold. (d) The wafer is
coated with an additional 160 nm of LPCVD low stress nitride to define the
pyramidal tip in silicon nitride. (e) A 650 µm square hole is etched in the
backside silicon nitride as in (b). (f) The silicon is etched by immersion in 30%
KOH in water at 85 °C to release the membrane................................. 62

Figure 4.3 – (a) C-NSOM scan of chrome nano-wires oriented parallel to the
polarization. (b) The corresponding AFM image for (a). (c) C-NSOM scan of
chrome nano-wires oriented perpendicular to the polarization. Debris obstructs
the center of the 130nm nano-wire; the right hand side is unobstructed. (d) The
corresponding AFM image for (c)................................................................ 64

Figure 4.4 – (a) C-NSOM line scans across TM Chrome nano-wires. Data is offset
vertically by 0.15, 0.1, 0.05 and 0.0 for the 50, 85, 130 and 235 nm wires
respectively. (b) C-NSOM line scans across TE chrome nano-wires. Data offset
vertically as in (a). Arrows indicate the positions of (c-h). (c,f) Diagram (c) and
FDTD simulation of normalized charge density (f) of a cross section along the
red dashed line of Fig. 1(a) showing the charge displacement in the C-aperture
during normal operation. Illumination is from the top. Emission to the far-field is
shown as red arrows. (d,g) Diagram and simulation of operation with a chrome
wire bridging the gap. (e,h) Diagram and simulation of operation with a chrome
wire offset away from the gap. ........................................................................ 66

Figure 4.5 – Fabrication of chrome step edge. (a) Evaporation of water leaves an
aggregate of polystyrene spheres. (b) Evaporation of chrome. (c) Removal of the
polystyrene spheres and any chrome on top. (d) The liftoff leaves protrusions
along some of the edges..................................................................................... 67

Figure 4.6 –ERe) The intensity profile of the C-NSOM probe tip. Transmission of the
illumination spot through the gold layer produces a large background envelope
(mode field diameter of 2.1 µm). The small near-field spot produced by the C-
aperture is superimposed on the wider leakage background. (b) The transmission
profile that results from scanning over a step edge; essentially a convolution of
the curve in (a) with a step. (c) Experimental observation of the center portion of the (b) with a fit to the data using Equation (4.1).................................68

Figure 4.7 – (a) C-NSOM image of chrome step edges. (b) AFM image of the same structure as (b). Note the varying height of individual protrusions along the chrome edge.................................................................70

Figure 4.8 – Sigmoid widths for fit to C-NSOM line scans over chrome edges with protrusions of varying heights (black). Linear fit to the sigmoid widths (red). Simulated FWHM along the x axis plus a 101 nm offset (blue).................................71

Figure 5.1– C-NSOM setup. (a) The aperture shape used in this study, defined by the characteristic size, $\alpha$. (b) FDTD simulated near-field spot cross sections normalized to $|E_{inc}|^2$ in the center of the a-Ge$_2$Sb$_2$Te$_5$ layer. (c) An SEM image of an FIB cross sectioned pyramidal tip corresponding to the dotted box area in (d). Tilt angle is 52°. (d) The assembled C-NSOM probe tip shown oriented above the recording medium (not to scale). Light is delivered from the top. The recording medium consists of (from top to bottom): a silicon nitride capping layer to prevent Ge$_2$Sb$_2$Te$_5$ oxidation, the a-Ge$_2$Sb$_2$Te$_5$ recording layer (shown with c-Ge$_2$Sb$_2$Te$_5$ recorded bits), a silicon nitride thermal buffer layer and a titanium heat sink layer...........................................................................77

Figure 5.2 – Finite Element Simulation of Temperature. The temperature in the center of the Ge$_2$Sb$_2$Te$_5$ layer simulated using COMSOL Multiphysics for pulse lengths of 100 ns and 10 $\mu$s with far-field illumination spot sizes (1/e$^2$) of 1.0 $\mu$m and 2.1 $\mu$m respectively. The 10 $\mu$s pulse power is set at the threshold power and illumination spot size from experiment. For the 100 ns pulse, the far-field illumination spot size has been reduced to a value attainable with an NA = 0.6 lens (the same as used in DVD) and the power set to produce the same temperature as the 10 $\mu$s pulse.................................................80

Figure 5.3 – Optical and AFM bit readout. (a) A transmission C-NSOM image of bits recorded with 10 $\mu$s, 19.34 mW pulses. (b,c) Cross sections along the white dashed line in (a). FDTD simulated C-NSOM cross sections for an elliptical crystalline bit with x-y dimensions of 60x50 nm are shown in as circles. A halo of increased transmission can be seen around the transmission dip in (a-c).............81
Figure 5.4 – Surface topography after recording. AFM data of the capping layer that was simultaneously acquired with the C-NSOM scan shown in Figure 5.3a. The surface is flat to within ~4 nm, demonstrating that the transmission changes measured in Figure 5.3a are optical in origin. Additionally, because future near-field recording systems must incorporate read/write flying heads similar to those found in today’s hard drives and operating fly heights on the order of 10 nanometers, it is extremely important that there is no surface deformation after writing to prevent head crashes.

Figure 5.6 – (a) An AFM scan of bits recorded with 10 µs, 19.34 mW pulses. White arrows and dashed cross hairs indicate the position of recorded bits. (b,c) Cross sections along the bit identified in (d) by the cross hairs. Dashed lines show a two-sided logistic fit to the bit depression.

Figure 5.7 – Bit FWHM versus Optical Power. (a) The average bit size as measured by AFM and the average measured optical response FWHM for bits recorded with t_pulse = 10 µs and varying optical power. Dashed lines are FDTD simulated C-NSOM optical response FWHM for bits with x-y dimensions of 60x50, 80x60 and 90x70 nm (chosen to match the measured AFM values). Error bars are 1σ. Images (b) and (c) show example C-NSOM images of bits recorded at low and high power, indicated by the dotted ovals in (a). Images (d) and (e) show examples of the corresponding AFM images. Note the change of scale.

Figure 7.1 – Example cross section of the simulation space taken along the x axis and through the aperture ridge. This simulation happens to be from Chapter 5 and is of a 60x50 nm crystalline bit slightly offset in x from the aperture center.

Figure 7.2 – The height map for an α = 50 nm aperture used for “milling” the aperture in the simulation.

Figure 7.3 – Example plot of the electric field amplitude at the aperture exit versus time.

Figure 7.4 – An example RCS pattern for an α = 50 nm aperture over a-Ge2Sb2Te5.

Figure 7.5 – The C-NSOM setup showing the components external to the tip sample interface. Note that the setup shown here is rotated 90° counter-clockwise from the inverted microscope configuration of the experiment.
Figure 7.6 – Z-axis flexure stage. The piezo stack is inserted in the u-shaped gap forcing the open box to deform and translate down. The tuning fork is mounted at the end of this structure. .......................................................... 102

Figure 7.7 – (a) A droplet of glue is placed on the end of the cleaved fiber facet. (b) The fiber tip is touched to the membrane frame surface away from a membrane and the glue moves to fill any gaps. (c) After UV curing the glue, the tip and glue are snapped free of the surface. (d) A very small amount of additional glue is added to the tip, the fiber core and pyramidal tip are aligned and the glue is UV cured. (e) The membrane is broken free from the silicon frame. ........................................... 104

Figure 7.8 – (a) Tuning fork oscillation amplitude versus frequency. (b) Tuning fork oscillation amplitude versus tip-sample distance. ............................................................. 105
1 Introduction

The limitations of optical imaging using lenses have long been recognized. Most important today are the physical limitations imposed by the nature of light, specifically the diffraction limit. The diffraction limit (or Abbe limit) restricts the length scales on which propagating light can distinguish separate physical objects to a size \( d \), where

\[
d = \frac{\lambda}{2NA}
\]  

(1.1)

and where \( \lambda \) is the wavelength of light, and \( NA \) is the numerical aperture of the focusing lens \( (NA = nsin(\theta), \ n \) is the local refractive index, and \( \theta \) is the angle between the marginal ray and the optical axis). It has ramifications for many fields of science and engineering, two of which are studied in detail in this dissertation: microscopic imaging and optical data storage.

Microscopic imaging is significantly hindered by the diffraction limit, which puts the imaging of many objects of interest outside the capabilities of the optical microscope. It would, for instance, be very desirable to directly photograph DNA and other large molecules, or some of the many nanoscale structures that are being fabricated and studied today. Unfortunately, optical microscopes that operate in or near the visible regime are capable of image resolution of only \( \sim 200 \text{ nm} \) \( (\lambda = 400 \text{ nm}, \ NA = 1) \) and with oil immersion objectives \( (NA = 1.6) \) or solid immersion lenses (SIL) \( \sim 63 \text{ nm} \). As an alternative to increasing the numerical aperture, the wavelength can be reduced, but, as the semiconductor industry has discovered, the light sources and optics required to focus light with a wavelength shorter than
200 nm are expensive and present a host of new problems such as large absorption in air and other materials.

When, instead of collecting light scattered from an object, we are interested in focusing light onto an object to affect some change (e.g., local heating as in the case of optical data storage) the same approaches of increasing \( NA \) and decreasing \( \lambda \) can be employed, but the diffraction limit problem remains. In the field of optical data storage, this difficulty can be clearly seen in the evolution of the employed wavelength and numerical aperture from compact disks (\( \lambda=780 \) nm, \( NA=0.50 \)) to Blu-ray (\( \lambda=405 \) nm, \( NA=0.85 \)) and the apparent lack of options for further increasing data density. Immersion techniques (in the form of SILs) have been tested for optical data storage applications but with limited success (see Chapter 5 - C-Aperture Based Optical Data Storage). Holographic techniques have also been extensively explored and although they are an extremely promising approach to the problems facing optical data storage, they still do not fundamentally overcome the diffraction limit.

### 1.1 The Optical Near-Field and the NSOM

In order to transcend the diffraction limit, we must turn to near-field techniques. To understand why, let us consider the generalized electric field distribution \( U(x,y,z) \) at a point \( P(x,y,z) \) where the field consists of an amplitude and a propagation vector \( k(k_x,k_y,k_z) \). The magnitude of \( k \) is given by

\[
|k| = \frac{2\pi n}{\lambda}
\]  

(1.2)

We wish to determine how close together two adjacent points (\( P_1 \) and \( P_2 \)) can be while remaining measurably distinct. Our measurement of the field at point \( P \) has some uncertainty in position (\( \Delta x, \Delta y, \Delta z \)) and in the propagation vector (\( \Delta k_x, \Delta k_y, \Delta k_z \)) and the Heisenberg uncertainty principle states that (limiting ourselves to the x dimension) [1, 2]
\[ \Delta x \Delta k_x \geq 2\pi \]  

(1.3)

and noting that for the minimum \( \Delta x \) value, \( \Delta k_x \) yields its maximum value by spanning the available momentum space.

\[ \Delta k_x = 2k_x \]  

(1.4)

We can express \( k_x \) as a projection along \( k \) as

\[ k_x = |k| \sin \theta \]  

(1.5)

and combining Eq. (1.2), Eq. (1.4) and Eq. (1.5) to rewrite Eq. (1.3) we have

\[ \Delta x \geq \frac{\lambda}{2n \sin(\theta)} \]  

(1.6)

which is simply a restatement of the Abbe limit in Eq. (1.1). For propagating waves, \( k \) has real values and \( n \) is limited to the available physical values of at most 2-3 and we are limited by the diffraction limit. However, if we allow \( k \) to assume complex values, i.e., non-propagating or decaying waves, then instead of satisfying Eq. (1.5), the modulus of \( k \) must equal the sum of the squares of its components.

\[ |k_x| = \sqrt{|k|^2 - (k_y)^2 - (k_z)^2} \]  

(1.7)

If \( k_y \) or \( k_z \) are complex, then \( k_x \) can exceed \( |k| \) in magnitude which allows values of \( \Delta x \) smaller than in Eq. (1.6) and it is this quality that allows us to access information about subwavelength optical fields but only at external measurement points \( P \) that are within the range of the decaying fields generated at point \( P_1 \).

Subwavelength dielectric objects necessarily produce imaginary scattered components because of the high spatial frequencies of their surfaces. This is also true of metallic objects with the additional opportunity to obtain imaginary \( k \) vectors from surface plasmon sources.
The classic example of this is the flat metal/dielectric interface where surface plasmons can propagate with $k_x$ and $k_y$ real and $k_z$ imaginary (see §1.3 - The Apertureless NSOM).

Figure 1.1 – (a) Conceptual NSOM design operating in illumination-transmission mode. (b) Operating in collection mode. (c) Operating in illumination-reflection mode.

A method to access the information present in the near-field was first conceived by Synge in 1928 [3] (later to be developed into the near-field scanning optical microscope (NSOM)). The idea in its simplest incarnation consists of a hole in an opaque screen adjacent to the object of interest. Figure 1.1a shows a diagram of the Synge microscope where a light source illuminates the aperture with a width much smaller than the wavelength of the incident light. The aperture plane is held very close (relative to the illumination wavelength) and translated with respect to the sample. In this diagram, optical detection is in transmission mode and collection in the far-field.

Introduction of the light can be achieved in several configurations: illumination-transmission mode (shown in Figure 1.1a), collection mode (Figure 1.1b - the sample is illuminated from the far-field and light collected through the aperture constitutes the signal), and illumination-reflection mode (Figure 1.1c - light reflected by the sample and transmitted back through the aperture is collected and constitutes the signal). The illumination and collection modes have been shown to be equivalent [4].

Unfortunately for Synge and his contemporaries, the technological means of implementing an NSOM were not realized until 1972 when an instrument was demonstrated.
in the microwave by Ash & Nicholls [5]. An instrument of more general interest (operating in the optical regime) did not appear until 1984 when A. Lewis et al. and D.W. Pohl et al. independently invented the tapered optical fiber NSOM [6, 7]. This instrument remains today the most common implementation of Synge’s idea. One of the most significant technological obstacles to overcome to produce the NSOM was the fabrication of apertures sufficiently small to be of interest to imaging in the nano-optical regime.

The problem of fabricating aperture probes was solved using optical fibers that were heated and pulled to a fine point (an example is shown in Figure 1.2a). Heating is accomplished either with a direct flame, electric arc or CO2 laser. The sharp tip is coated with aluminum as shown in Figure 1.2c such that a small opening remains at the end (Figure 1.2b). Alternatives to heating and pulling such as chemical etching have been developed and have been shown to be capable of producing sharp tips with as much as 3 orders of magnitude higher transmission efficiency by fabricating tips with multiple tapers [8, 9]. The fabrication of the aperture has also been approved upon by using focused ion beam (FIB) milling of the aperture to gain fine control over the aperture size and produce smooth probe ends free of metal grains [10, 11].

Figure 1.2 – (a) An early NSOM probe fabricated by heating and pulling an optical fiber. (b) Aperture at the end of the fiber produced by rotation of a pulled fiber during Al vapor deposition. (c) Configuration used for coating optical fibers with Al to produce a small, round aperture at the tip. Images reproduced from [12].
1.2 Limitations of Round Apertures

In all of the preceding discussion, the optical near-field has passed through a small round aperture to force localization and I have made no mention of the efficiency with which light is transmitted through these apertures. In fact, the transmission efficiency through round (and square) deeply subwavelength apertures is a significant problem in near-field optics. In 1944, Hans Bethe in his famous paper “Theory of Diffraction by Small Holes” [13] showed that when the size of a hole, with radius $a$, in a perfectly conducting infinitesimally thin screen was much smaller than the wavelength, the power transmission drops precipitously. He found that the transmission cross section $\sigma_t$ in the far field is

$$\sigma_t = \frac{\int |S| r^2 \sin \theta d\theta d\phi}{S_i} = \frac{P_{trans}}{S_i} \propto a^2 \left( \frac{a}{\lambda} \right)^4$$  \hspace{1cm} (1.8)

where the integral over the Poynting vector $S$ includes the aperture exit half space yielding total transmitted power $P_{trans}$, the incident power density is $S_i$, and the wavelength of incident light is $\lambda$. We can see from this result that for apertures much smaller than the wavelength of light there is virtually no power transmitted. Unfortunately, these very small apertures are exactly the kind we are interested in because of our desire to break the diffraction limit by much more than a few factors.

Figure 1.3 – A plot of power throughput vs. aperture radius for $\lambda = 1 \mu m$. Dots are finite difference time domain simulations of round apertures in an infinitesimally thin perfectly conducting screen. Solid line is the Bethe power law. Plot reproduced from reference [14].
Power throughput, $PT$, was defined by Shi [14] as the transmission cross section $\sigma_t$ normalized to aperture area and we can see in Figure 1.3 that at $\lambda = 1 \mu m$, even relatively large holes with radius 100 nm have optical transmission more than 2 orders of magnitude lower than the incident field with the sharp drop in transmission starting at ~250 nm ($\lambda/4$).

The typical implementation of an aperture fiber probe tip yields a transmission efficiency of $10^{-4}$ due both to the low aperture transmission and the elimination of propagating modes in the fiber taper before the light even reaches the aperture. Although this low power is acceptable for laboratory near-field imaging applications where sensitive detectors, long integration times and slow image acquisition rates are permissible, it is catastrophic for more demanding imaging applications.

In the case of optical data storage where high delivered power is critical the situation is hopeless. Spot sizes must be $\leq 100$nm to be of interest in optical data storage (this value improves on Blu-ray by about a factor of 2 in linear density or 4 in areal density) and Figure 1.3 shows that apertures of this size produce power throughputs $\sim 3.2$ orders of magnitude lower than the incident light. Given that optical data storage systems typically use recording powers of $\sim 10$ mW, we would need more than 10 W of power to record data using a 50 nm radius aperture! Finite difference time domain (FDTD) simulations bear this out (see §7.1 - Finite Difference Time Domain (FDTD) Simulations) as we can see in Figure 1.4. The aperture shown is a 25 nm radius circular aperture in a 120 nm gold film sourced by a plane wave. The image is a cross section through a plane parallel to the metal film and displaced from the exit surface by 30 nm. Plotted is the magnitude squared electric field normalized to the magnitude square electric field of the incident light. The transmitted intensity is even worse than predicted in Figure 1.3 ($4.1 \times 10^{-5}$ vs. $\sim 3 \times 10^{-4}$) because of the extra distance between the aperture exit and the data storage medium as well as the thick real metal film.
Figure 1.4 – FDTD simulation of a 50 nm round aperture near an optical data storage medium. Side lobes are due to leakage of the illumination light at the edges of the Drude metal and perfectly matched layer (PML).

1.3 The Apertureless NSOM

The most common alternative to the significantly flawed subwavelength aperture is a localized near-field spot produced by illuminating a sharpened tip with a far-field beam of the correct orientation and polarization. During operation, the illuminated tip acts as a near-field light source and the sample scatters that near-field into the far-field for detection. This method is referred to as apertureless NSOM or ANSOM, first demonstrated in 1994 [15, 16]. Amazingly, the idea was also suggested by Synge in a letter to Einstein but was never published because of concerns that the local tip-to-sample interaction would be overwhelmed by the signal from the irradiating light [17]. The sharpened tip is usually a metallic or semiconductor material and often piggybacks on the technology developed for atomic force microscopy (AFM) using cantilevers with chemically etched tips. The tip is illuminated as shown in Figure 1.5 with the electric field polarized along the axis of the tip, forcing the free electrons in the tip to oscillate with the optical field. At the apogee of each oscillation, electrons/holes bunch into the tip creating an extremely high electric field. This is often called the “lightning rod effect” because of the way in which electrons are forced to flow along the
surface to the desired area and it will be revisited in §1.4.2 as the primary method by which bowtie, H, C-apertures and many other shaped apertures achieve small near-field spots.

The field intensity at the end of the tip is typically modeled to be $10$ to $10^5$ times higher than the incident field depending on the tip optimization and the local environment [18-20]. Additionally, sharpened tips have been experimentally demonstrated to produce an electric field that is an order of magnitude higher in intensity if there are sharp conducting surface features placed opposite the tip [21]. “Tips” that consist of a single silver nanoparticle have been shown to produce Raman enhancement factors of $10^{14}$ to $10^{15}$ [22]. The resolving power of ANSOM is extremely good because the spot size is determined predominantly by the radius of the tip and in the infrared, resolutions of $\lambda/600$ have been demonstrated [23]. In the visible, single molecules can be imaged using ultra-sharp SiO$_x$ whiskers [24].

![Figure 1.5 – Apertureless NSOM configuration.](image)

The very high local field intensity and excellent resolution make the ANSOM seem like an ideal solution to sub-diffraction limit imaging and data storage. Unfortunately, because the illumination light is present over a much larger area compared with the near-field light at the tip, the background signal therefore integrates to a larger signal than the tip enhanced signal and picking the super-resolution signal out of the background can be difficult. This problem is one of the greatest drawbacks of the apertureless approach. Commonly, the
height of the tip is modulated and lock-in techniques are used to isolate the signal coming from the near-field source [25] and although this leads to much better signal to noise ratios, it necessarily slows down scanning, making the method inappropriate for high throughput imaging or data storage applications where the data readout rate is very important.

1.4 Enhanced Transmission Apertures

Fortunately, in the last decade, a slew of innovative aperture designs have been proposed that improve on the low performance of round or square shaped apertures and often incorporate the lightning rod effect in conjunction with a propagating (or near-propagating) waveguide-like mode to give the best of both the apertured and apertureless worlds. Generally, enhanced transmission apertures fall into two categories that will be discussed in the following sections. The first relies on coupling of surface plasmons (SPs, described in detail in §1.4.1 - Aperture Arrays) by a structure surrounding a single central aperture. The structure consists of a 1D or 2D array with a grating period appropriate for coupling and launching SPs along the surface towards the central aperture. The second category (§1.4.2 - Transmission Mode Apertures) consists of single apertures possessing a propagating mode that allows for efficient input and output coupling and does not rely on SPs.

1.4.1 Aperture Arrays

The field of enhanced transmission aperture arrays was launched in 1998 when Ebbessen published a paper describing transmission greater than unity for aperture arrays where the apertures were much smaller than the wavelength of light [26, 27]. They attributed the high transmission to coupling of the illumination light to SP modes that were then re-emitted on the exit surface of the metal. Surface plasmons are non-radiative collective electron oscillations bound to a metal/dielectric interface having real in-plane wave vectors ($k_x$,
and $k_\parallel$) and an imaginary out of plane wave vector ($k_z$). The relative permittivities have the requirements that

\begin{align}
\epsilon'_m < 0 \\
\epsilon''_m < |\epsilon'_d| \\
|\epsilon'_m| > \epsilon_d
\end{align}

(1.9) (1.10) (1.11)

where the relative permittivity of the metal is $\epsilon_m = \epsilon'_m + i\epsilon''_m$, and $\epsilon_d$ is the relative permittivity of the dielectric. The derivation and physics have been reviewed in numerous places so I will not bore the reader with a full rehashing; a canonical explanation is available in Raether’s text *Surface Plasmons on Smooth and Rough Surfaces and on Gratings* [28]. Aside from the permittivity requirements in Equations (1.9), (1.10) and (1.11), the important end result is the dispersion relation for the in-plane wave vector:

\[ k_{sp} = \frac{\omega}{c} \sqrt{\frac{\epsilon'_m \epsilon_d}{\epsilon'_m + \epsilon_d}} \tag{1.12} \]

where $k_{sp}$ is the wave vector of the surface plasmon, $\omega$ is the free space frequency and $c$ is the speed of light. The dispersion curve in Eq. (1.12) is normally inaccessible by light propagating in free space and special configurations such as Kretschmann coupling or grating coupling [28] are used for the coupling of far-field light into SPs. There also exists a category of SPs called localized surface plasmons (LSPs) that are found propagating on the surface of nanometer scale metallic structures (the common example is gold nano-particles which exhibit absorption resonances in the visible and can be used to make colored glass). LSPs are excited by adding the momentum from whatever grating vector is present in the Fourier decomposition of the metallic structure’s physical shape. In the case of aperture arrays, the periodic structure defines a two dimensional grating which provides the coupling to the
surface plasmon mode and also provides the pathway by which evanescent light on the exit surface is out-coupled to propagating modes. There will then be transmission maxima when

\[ k_{sp} = k_x \pm nG_x \pm mG_y \]  

(1.13)

where \( n \) and \( m \) are integers, the grating vectors are \( G_x = G_z = 2\pi/a_0 \), and \( a_0 \) is the grating period. Generally the in-plane wave vector \( k_x \) is zero and the grating vectors must match the momentum of the surface plasmon.

Although this result is very interesting, it is not very useful for application to near-field inspection or data storage because the light is emitted from all the apertures in the array instead of from a single point source. Fortunately, it was not long after the discovery of enhanced transmission aperture arrays before the configuration was modified to become a dimple array or concentric ring corrugations with a single through aperture in the center of the array [29]. This configuration leads to similar results where the transmission normalized to the aperture area and measured in the far-field is on the order of unity. Because momentum matching to flat gratings with normally incident light launches SPs both towards and away from the central aperture in equal parts, more recent implementations use a tapered cone to set the angle of incidence and to preferentially launch SPs towards the aperture with increases of about 10x over classical tapered fiber tips [30]. Tips that have a broad taper angle (that is, nearly flat) are modeled to produce local intensity enhancements over the incident light of around 36 times, although for an aperture size \( \sim \lambda/4 \) [31], which is quite large and when compared to Figure 1.3 can still be seen to represent an increase of only about 10x over the expected single aperture result.

Ultimately, using periodic structures to enhance the throughput of round apertures is not an ideal approach. In addition to the modest intensity improvement of 10x, the periodic structure must be large (on the order of 10\( \lambda \)) to efficiently couple [32] and this means that a
very large far-field optical spot must be used to illuminate the device. When compared to
illuminating a single nano-aperture with a nearly diffraction limited spot, the total power
required (and thus the local heating) is ~100x larger, which is detrimental to many
applications. In our data storage example of the 50 nm radius aperture from §1.2 we would
need an intensity 10x less than before because of the grating enhancement, but the
illumination area is 100x larger requiring 10x the total power or 100 W!

1.4.2 Transmission Mode Apertures

At about the same time as the Ebessen discovery of enhanced transmission periodic
arrays, there appeared in the literature studies using single apertures and antennas illuminated
with near-diffraction limited spots that produced extremely high near-field intensities. These
structures were chosen for their high scattering cross section to maximize their interaction
with far-field light, hopefully coupling that light into a tightly confined near-field spot.

Today, there are two fiefdoms: groups studying bowtie apertures/antennas and groups
studying C or H-apertures. Although different groups often concentrate on and promote one
design over all others, they are functionally similar (particularly bowties apertures and H-
apertures) and can be understood as one design with different dimensions that improve or
detract from their performance in a particular application. This can be clearly seen when
examining the field distribution of the lowest order mode and the charge displacement in
bowtie and H apertures (see Figure 1.10). In both cases, field lines connect the ridges/tips and
the ridges/tips also form a dipole. The C-aperture can be viewed as an H-aperture with a
ground plane inserted in the center (as has long been understood in the microwave regime) and
the analog of the C-aperture bowtie is a C-aperture with a sharpened ridge (again, see Figure
1.10). The surface currents and field distributions of bowties, C and H-apertures will be
discussed in more detail in §1.4.2B - Bowties and §1.4.2C - C and H-Apertures.
A. Aperture and Antenna Fabrication

Presently, there are two primary and competing methods for making nano-apertures and antennae: electron-beam (e-beam) lithography and focused ion beam (FIB) milling. In e-beam lithography (Figure 1.6a-c) a thin layer of e-beam resist is spun onto a conducting substrate – usually poly (methyl methacrylate) (PMMA) spun onto a fused silica substrate which has had ~5 nm of indium tin oxide (ITO) deposited and sintered on top. The spun on layer of resist is exposed with an electron beam in the areas where metal is to remain and developed such that the resist edges are undercut (Figure 1.6a). Next, the metal of choice is deposited on the entire surface (Figure 1.6b) and to avoid connections between the metal deposited on top of the resist and the metal deposited directly on the substrate, evaporation is typically used because of the good collimation of the metal ions in the vapor. Finally, in the liftoff process, the resist is removed by a solvent, leaving metal only where desired (Figure 1.6c). The primary benefit of this technique is that the metal that is deposited is of very high quality and does not need to be exposed to anything but organic solvents before use as an optical device. Unfortunately, e-beam lithography is much better suited to deposition of small metal structures (antennae) rather than the creation of holes. This is especially true when working at the limits of resolution (~20 nm), where the required undercut is unimportant for antennae but increases the feature size of apertures.

The other commonly used method of nano-structure fabrication is FIB milling (Figure 1.6d&e). In this method, apertures are made by directing a beam of tightly focused ions onto the sample surface and raster scanning the beam in the desired pattern. Ions incident on the surface collide with surface atoms, ejecting material through a physical sputtering process and leaving a hole. Gallium ions are almost always used although there are a few helium ion beam tools in the world. The primary advantage of this technique is that the instrument operator directly fabricates the apertures with no intermediate steps where errors can be introduced.
Beam current, spot size and dwell time are the primary factors that affect the resulting hole and feedback on the quality of the hole produced is readily available since most FIB instruments also incorporate an SEM for high resolution inspection. Obviously, this method is not well suited to fabrication of antennae since it takes an inordinate amount of time to remove all the surrounding material. However, the biggest drawback of FIB milling is the surface damage that results; ion implantation as well as material mixing from redeposition are significant problems. Also, because the ion beam has a finite spot size, the edges and corners of the resulting structure are not perfect and can significantly affect the performance of the aperture (this is the subject of Chapter 2). Although still a problem, round edges from a finite beam size are not as troublesome in e-beam lithography because resist exposure and development are together a nonlinear process.

Figure 1.6 – E-beam vs. FIB nano-structure fabrication. (a) e-beam lithography of a resist. (b) Metal deposition. (c) Lift-off of the unwanted metal. (d) Physical etching using a gallium ion beam. (e) The resulting pit with ion damage and material mixing.
B. Bowties

The next sections will address two of the most common apertures/antennae starting with bowties. The first proposal and demonstration in the microwave regime of a bowtie antenna was published in 1997 by Grober et al. [33]. Numerical studies in the optical regime followed for antennae [34] and apertures [35] that showed that bowties are strongly resonant structures and that when operated at the correct wavelength can have near-field intensities ~10x larger than circular apertures with equivalent near-field spot size. The wavelength of maximum transmission is conveniently located in the visible for bowties that are about 200 nm in height (see Figure 1.7a) and having gaps of about 17 nm [36].

![Figure 1.7 – Bowtie configurations: (a) Bowtie antenna. Polarization of the illuminating field is vertical. (b) Bowtie aperture. Polarization of the illuminating field is horizontal. The gap between bowtie tips is g and the radius of curvature of the tips is r. +/- signs represent displaced charge in the metal during operation. Dotted arrows indicate the direction of surface currents.](image)

Diagrams of a bowtie antenna and aperture are shown in Figure 1.7. The antenna consists of two triangles, the tips of which have some non-zero radius of curvature, r, and the triangles are separated from each other by a gap g. During illumination, LSPs are generated that move along the surface and generate a Hertzian dipole across the gap as indicated by the +/- signs in Figure 1.7a. The dipole oscillates at optical frequencies and the high charge density at the tips (another instance of the lightning rod effect) produces a very intense near-field spot. Because the net charge on the bowtie is zero, this also leads to a charge
accumulation on the flat surfaces of the bowtie that can act as secondary optical sources. The size of the individual triangles, d, in conjunction with the total height, h, determines the wavelength resonance condition for the bowtie.

The inverse case of the bowtie antenna is the bowtie aperture shown in Figure 1.7b. The aperture must be excited with light polarized 90° to that used for the antenna, which excites surface currents on the inner perimeter of the aperture [37]. The charge in these currents sloshes back and forth with the optical field and at each oscillation charge is bunched into the tips at the center of the aperture forming a dipole that utilizes the lightning rod effect to boost intensity. The surface currents in the aperture are not predominantly excited by SPs, since they exist in simulations that use perfect electric conductors (PEC; PECs do not support surface plasmons as a result of the total exclusion of field from their interior [38]). Instead, they are a result of a waveguide-like tunneling mode that will be discussed in more detail in §1.4.2C.

Experimental work has proved out the numerical predictions, first with demonstrations of the far-field resonant characteristics of bowties [36], followed by NSOM measurements of the near-field spot size [39]. Photoresist exposures demonstrating the tight optical confinement to the central tips as well as the parasitic spots that occur at high optical powers at the flat extremities were also demonstrated [40]. The work in Reference [40] clearly showed that the bowtie aperture produces two near-field spots of equal intensity, one at each triangle tip. Although the spots merge when the field is measured more than a few nanometers from the bowtie surface, they may still be a problem in some applications (a single spot is certainly desirable for data storage).

Difficulty in making high quality bowtie apertures has forced most groups to use e-beam fabricated bowtie antennas with all the background problems of ANSOM. The preferred method for manufacturing bowtie apertures is FIB milling [39] which damages the metal tips.
in the bowtie center. In fact, resonant nano-structures are so sensitive to material properties that Reference [36] notes that simple SEM imaging significantly alters the bowtie performance, either through e-beam induced changes in the crystalline structure of the metal or through e-beam assisted deposition of organics. FIB milling is even more destructive than SEM and should be expected to degrade the optical performance of milled metals. In addition to materials problems, fabricating sharp tips ($r < 20 \text{ nm}$) that maximize intensity and minimize spot size is very difficult when the FIB spot size is typically 30 nm or more.

C. C and H-Apertures

I now turn to the aperture that has been the primary focus of my research and to a discussion of the past work done in the conception, optimization and demonstration of the C-aperture’s efficacy. This section will cover the mode structure of the ridge waveguide and its relation to the C-aperture as well as the resonant transmission properties and the input and output properties of the C-aperture.

Although there is an important distinction between waveguides and apertures (see §1.4.2C Waveguide vs. Aperture), a microwave regime sibling of the C-aperture has actually been in use for a long time in the form the ridge waveguide [41] and diagrams of the single and double ridge waveguides (analogous to C and H apertures, respectively) are shown in Figure 1.8.

![Figure 1.8 – (a) Diagram of the C-aperture (single ridge waveguide). (b) Diagram of the H (or I) aperture (double ridge waveguide).](image-url)
Mode Structure and Optical Confinement

The ridge waveguide was developed because of the relatively wide separation between the primary mode (TE\textsubscript{10}) and the next higher order mode and because of the intermediate impedance of the waveguide; both reasons that are not of much importance to the apertures used in the optical regime (although the impedance affects coupling) [42].

![Figure 1.9 – Plot of the waveguide cutoff wavelength normalized to a (see Figure 1.8). The variables a, s, d and b refer to the dimensions in Figure 1.8. Image reproduced from Ref [43].](image)

Because of the folded structure of the ridge waveguide, the physical dimensions of the guide cross section in relation to the cutoff wavelength are substantially smaller than for the rectangular (or round) case. This can be seen in Figure 1.9 where the largest dimension, a, of the waveguide is shown to be 2.6-5.4 times smaller than the cutoff wavelength (c.f. rectangular waveguide with size $\lambda_c/2$). Already this is an improvement if our goal is to use the
exit of a truncated waveguide as a small optical source, but a much more substantial reduction in size is available if we look at the distribution of the electrical field inside the waveguide.

The appeal of using the C or H for near-field optical confinement can be seen in Figure 1.10 where the electric fields inside a selection of ridge waveguide shapes are shown. The figures were calculated using the 2D electromagnetic wave eigenmode solver in Comsol 3.2a and show the x-y electric field as arrows and the magnitude squared sum of the E_x, E_y, and E_z field components as the colormap. The walls are PEC so the result represents the field inside an infinite metallic waveguide in the microwave regime. Figure 1.10a shows a typical single ridge waveguide and the high electric field that develops between the capacitive surfaces of the ridge tip and the opposite interior wall. Immediately after the exit of a waveguide of this type (at distances \(<< \lambda/2\)), the emitted light will be concentrated where the electric field is highest, \(i.e.,\) in the gap region and at the tips of the ridge. If the two corners of the ridge are pulled together to form a tip, as in Figure 1.10b, then the electric field can be confined even more tightly to the area just surrounding the tip by further exploiting the lightning rod effect. However, in a structure that has a minimum feature size < 75 nm, it is much more likely that sharp features will be smoothed out during fabrication (with an FIB, for instance). In this case, the ridge will be rounded (Figure 1.10c) and the electric field intensity will be distributed more evenly throughout the gap region. Figure 1.10d&e show the fields for double ridge waveguides with both square and pointed ridges. This configuration is slightly less desirable for inspection and data storage applications because the high field at the ridge ends produces two near-field spots at the waveguide/aperture exit face. Figure 1.10f shows a waveguide with a bowtie aperture cross section to demonstrate that the field distribution is essentially identical.

Figure 1.10c represents the key to reducing near-field spot size with the C-aperture: there exists a propagating waveguide mode near the operating wavelength to produce high
transmission while at the same time the highest electric field in the waveguide is tightly confined to the gap region. This confinement in the optical regime has been measured several times, first by Chen et al. using ANSOM techniques [44], then by Wang et al. by exposing photoresist with an off resonance aperture [45] and later by Jin et al. using direct NSOM measurements [46]. References [45] and [46] both include convincing evidence that the intensity produced by a C-aperture is much greater than squares or circles that produce a similar spot size.

Figure 1.10 – Electric field distributions in infinite PEC waveguides at $\lambda = 980$ nm (scale bar under (a)). The colormap shows electric field intensity and arrows trace electric field lines, showing the proportional electric x-y field. (a) The standard C1 aperture design from Ref [14]. (b) Same as (a) but with the ridge tips pulled together to form a tip. (c) Same as (a) but with the ridge corners rounded with a radius of curvature 10 nm. (d) The H-aperture from Ref [14]. (e) Same as (d) but with the ridge tips pulled together to form a tip. (f) Same as (e) but with the ridge base point removed to form a bowtie aperture.

Waveguide vs. Aperture

Until now I have been somewhat vague about the distinction between ridge waveguides and C-apertures. Aside from differences in the thickness of metal in which the
waveguide/aperture is fabricated, the primary distinction between C or H apertures and their waveguide counterparts is found in the operating regime. Whereas ridge waveguides are operated purely at the TE$_{10}$ mode, C-apertures are operated below the cutoff frequency (longer wavelength) of this primary mode and achieve their high transmission by taking advantage of the high density of states and the lossy modes beyond cutoff that tunnel through the aperture [47].

![Image of maximum transmitted intensity for PEC waveguides of varying length. Curves entering from the left are Fabry-Perot like higher order resonant modes. Image reproduced from Ref. [48].](image)

The transition from aperture to waveguide can be seen very clearly in Figure 1.11 where at short guide lengths (i.e., aperture regime) the peak transmission occurs to the red of the waveguide cutoff (at ~780 nm). As the length of the guide increases, these lossy modes carry a decreasing proportion of the power until, at very long guide lengths, the cutoff wavelength is clearly demarcated and most power is carried in the TE$_{10}$ mode.

It is worth noting here that, as with bowties and unlike in the case of periodic arrays or surrounding gratings, SPs are likely not a significant contributor to the transmission through the aperture. This is evident from the very good performance of C-apertures in PEC and the lack of any significant structure that could support surface modes [38]. It is conceivable that
surface modes might be generated across the d or b gaps of the apertures but this would require that the resonant aperture size is determined predominantly or entirely by these dimensions. References [49] and [50] show that you can in fact stretch the aperture either vertically or horizontally (changing d and b) and retain or improve the transmission resonance by adjusting the ridge shape, indicating that coupling into modes around $TE_{10}$ (both propagating and evanescent) is the primary transmission mechanism in C-apertures rather than coupling to SP modes.

![Diagram showing surface plasmon generation in nano-apertures.](image)

Figure 1.12 – Surface plasmon generation in nano-apertures. Grating matching occurs at the input and output edges making the corners the only source of SPs. (1) SPs propagating away from the aperture. (2) SPs propagating through the aperture. (3) SPs sourced at the aperture exit either by (2) or by a waveguide like mode and propagating away from the aperture.

Although surface plasmons do not seem to be the primary method of transmission through the aperture they are definitely present and their effects can be seen in the experiments detailed in Chapter 5. Figure 1.12 shows the three possible SPs around a single nano-aperture (reflections are omitted). The only SPs that contribute to total transmission are those in (2), which are generated at the input corners of the aperture as a result of any grating period present in the structure that allows momentum matching between the incoming far-field light and the SP.
**Resonant Transmission**

The cluster of modes above and below the TE\(_{10}\) cutoff when operating in the aperture regime produces one of the most salient features of the C-aperture which is the resonant transmission peak. The C shape was first proposed and studied as a near-field source in the optical regime by Xiaolei Shi of the Hesselink group in 2002 [51] and the transmission resonance has featured prominently in research since then. In early FDTD numerical studies, the aperture was placed in an infinitely thin PEC [52] and later included real metal parameters [49] but in both cases a strong wavelength resonance exists. Figure 1.13 shows an FDTD calculated example transmission resonance curve similar to horizontal slices in Figure 1.11 but calculated for a real metal (gold). The Fabry-Perot like resonances that were observed in Figure 1.11 can again be seen to the left of the main peak.

![Example transmission resonance curve for a C-aperture in a 100 nm thick gold layer on a 75 nm thick silicon nitride membrane. Aperture size parameters: a = 195 nm, b = 145 nm, d = 50 nm, s = 65 nm.](image)

FDTD simulations show that the transmitted intensity of light at the near-field spot is much greater than that found in round and square apertures of both equal area and equal spot size. For the C-aperture in PEC compared to an equal spot size square, the PT is 10\(^3\) times larger for the C-aperture [49]. When the near-field optical intensity is measured directly, the
C-aperture typically produces intensities 1-10^2 times higher than the incident light. This feature of increased intensity is extremely desirable for data storage applications and has been the driver for much of the research into C, H and bowtie apertures.

An important feature of the aperture resonance is that it is closely linked to the size (perimeter) of the aperture. If the aperture dimensions shown in Figure 1.8 are held constant relative to each other, as is the case in most of the work I describe in this dissertation, then a percentage change in aperture size (a,b,d and s scaled by the same factor) leads to the same percentage change in resonant wavelength. For example, if a specific aperture design is resonant at 850 nm and you desire an aperture designed for 1 µm, scaling all the dimensions by 1.0/0.85 will produce a good first estimate aperture design. The relationship breaks down near metallic interband transitions (e.g., near 500 nm in gold) and when the skin depth becomes significant in comparison to the aperture’s feature size (δ ≥ s/2, where s is the ridge width shown in Figure 1.8 [48]).

Many studies have examined both the wavelength shifts and intensity changes in the transmission resonance for differing surrounding media, particularly apertures near phase change or magnetic recording media [50, 53-57]. Because the TE_{10} mode is only excited by incident light polarized along the b dimension of Figure 1.8, the aperture resonance is also strongly dependent on polarization. In fact, for cross polarized excitation, the resonant transmission peak completely disappears. The presence of the transmission peak, its polarization dependence and the improvement in transmitted optical intensity over square and round apertures were demonstrated experimentally by Matteo et al. in 2004 [58]. The result is shown in Figure 1.14 where images of the far-field transmitted light from incident white light of horizontal and vertical polarizations show the shifting peak transmission wavelength of the C-aperture for horizontal polarization. Vertically polarized excitation light shows no shifting resonance and a much lower transmitted intensity compared to the horizontally polarized case.
Matteo also demonstrated that the refractive index at the aperture exit can substantially shift the position of the resonance peak position [59], a property that will become especially important in Chapters 4 & 5.

![Figure 1.14 – Far-field optical transmission through a variety of nano-apertures. Aperture shapes are shown in the left hand column. Incident light is white. (a) Transmission for horizontally polarized incident light. (b) Transmission for vertically polarized incident light. Image reproduced from [58].](image)

**Input and Output Coupling**

I have already mentioned that the coupling into the C-aperture cannot be predominantly regulated by far-field coupling to SPs that then propagate through the aperture and so the obvious question then arises: how is light so effectively coupled from the far-field into the aperture? Sun *et al.* have studied coupling into square, rectangular and C-apertures using vector field topology visualization tools to analyze the position of power flow lines and critical points [60]. They found that square subwavelength apertures that do not possess a propagating mode have a saddle point in the vector field at the center of the aperture that forces nearly all the incident light away from the aperture. This saddle point, and how high above the aperture plane it occurs, strongly affects the amount of light passed through the aperture. Generally, the further above the aperture plane the saddle occurs, the more light is transmitted. For resonant C-apertures, this saddle point also exists but is moved well away
from the aperture plane allowing power to flow around it and into the aperture. By observing the power flow lines, it can be seen that the aperture behaves as an optical funnel and collects light from an area of about $\lambda^2$, much larger than its physical cross section. Additionally, there are two attracting focus critical points located at the tips of the ridge on the entrance side and two repelling focus critical points at the tips on the exit side. These four foci are associated with coupling of light into the propagating mode of the aperture.

The behavior at the aperture exit is also important, particularly for the inspection and data storage applications that are the focus of this dissertation. I have already mentioned that the ridge acts to localize the electric field at the aperture exit and that this localization has been demonstrated experimentally. However, the evolution of the near-field spot as it propagates away from the aperture exit is critically important because small propagation distances (< 10 nm) can have a significant impact on the spot size and thus on the aperture’s effective resolution. On the exit side of the metal film, we most frequently care about the intensity and distribution of the electric field (near-field spot profile) in slices parallel to the metal plane. An example near-field spot profile immediately adjacent to the metal surface is shown in Figure 1.15a. Figure 1.15a shows an aperture in gold having a ridge that is rounded in the x-y plane. The concentrated field can be seen around the ridge tip just as in Figure 1.10 except that the field around the ridge penetrates into the metal because of the finite skin depth. The field produced by opposing charge collected on the left hand aperture wall can also be seen.

Figure 1.15 – Near-field spot profiles produced by a C-aperture in a 100 nm thick gold layer on a 75 nm thick silicon nitride membrane. Aperture size parameters: $a = 195$ nm, $b = 145$
nm, \( d = 50 \text{ nm}, \ s = 65 \text{ nm} \) (identical to Figure 1.13). (a) Spot profile immediately adjacent to the metal surface. (b) Spot profile 18 nm from the metal surface.

Figure 1.15b shows the same spot measured 18 nm from the metal surface. The spot has enlarged substantially, becoming much more round. As expected for an expanding spot, the intensity has also decreased from a maximum value of \(~430\) times the incident field to \(~68\) times the incident field. The decrease is due partly to the geometric increase in size and partly to the exponential decay away from the metal surface experienced by near-field light.

![Figure 1.16](image)

**Figure 1.16** – Spot evolution vs. propagation distance from the aperture exit for the same aperture as in Figure 1.13 and Figure 1.15. (a) \(x\) and \(y\) FWHM. (b) Spot maximum intensity.

Figure 1.16a plots the evolution of the FWHM of the near-field spot at increasing distance from the metal exit plane. Initially, the FWHM expands very rapidly but at a slowing rate until \(~40\) nm where the expansion becomes linear. The rate of expansion is substantially different when the exit medium refractive index is changed or when layers of differing refractive index are present, which is a fact that will be important in later chapters. Another interesting quantity in the spot evolution is the maximum intensity, shown in Figure 1.16b. Here we can see that the intensity drops very rapidly and non-exponentially, as expected. In just the first \(~10\) nm, the intensity drops by an order of magnitude but it takes \(~40\) nm for the intensity to drop another factor of ten. The rapid decrease is an indication of the importance
of placing the aperture as close as possible to any object that we would like the aperture to interact with.

**Summary of Important C-Aperture Properties**

It is worthwhile at this point to reiterate the properties of the C-aperture that will be important in later chapters of this dissertation. These properties are:

- The C-aperture is a resonant structure.
- When operated on resonance, the intensity at the aperture exit is ~1-100 times the incident intensity.
- The aperture collects light from an area roughly \((\lambda/2)^2\), funneling light through the aperture.
- The electric field inside the C-aperture is concentrated at the tip of the ridge.
- The lightning rod effect can be used to increase the intensity at the end of the ridge.
- The near-field spot expands and loses intensity rapidly after the aperture exit.

**1.5 My Contributions and Thesis Outline**

The previous work on C-apertures presented in this chapter has consisted predominantly of simulation studies of the aperture’s ability to produce a small, intense optical spot. Although revealing, the existing experimental data is rather sparse compared to the simulation work and has focused on far-field studies of aperture resonance transmission, exemplified by the work by Matteo *et al.* [58]. Exceptions include the measurements of spot confinement noted in the ‘Mode Structure and Optical Confinement’ section. Unfortunately, none of the existing work measures or demonstrates the C-aperture’s promised on-resonance enhanced transmission and thus it remains to be shown that the aperture is capable of affecting a change in materials at a subwavelength scale as is required for optical data storage. Additionally, although the near-field of the C-aperture has been mapped, the aperture itself has
never been used to image other structures; an obvious application for a bright, small optical spot. My contributions focus on demonstrations of the practical fabrication and uses of C-apertures for near-field imaging and optical data storage and include:

1. Recognition and description of the two primary problems facing FIB based fabrication of near-field apertures: edge/corner rounding and gallium contamination.

2. Introduction of FDTD simulation that realistically represents the edge/corner rounding of the true aperture shape by modeling the FIB milling process as a mill pattern and Gaussian profile focused ion beam convolution.

3. Proposal and demonstration of a silicon nitride membrane based fabrication method that prevents surface corner rounding and gallium contamination during the FIB fabrication of nano-apertures.

4. Demonstration of the improved (8.8x) transmission performance of through membrane milled apertures over direct metal milled apertures using far-field measurements.

5. Demonstration of the near-field optical confinement at a resonant C-aperture’s ridge tip using two-photon absorption in negative photoresist as a near-field probe.

6. Designed, built and tested a C-apertured NSOM (C-NSOM) and demonstrated the first ever imaging with a C-aperture using test samples of chrome nano-wires and step edges.

7. Demonstrated a C-NSOM resolution of 100x74 nm (≤λ/10).

8. First ever demonstration of optical data storage using C-apertures, successfully recording data bits with a size of ~λ/20. This is likely the first aperture based recording of data bits sized significantly smaller than λ/4 in a phase change media.
The outline of this thesis closely follows the preceding list of contributions. I start in Chapter 2 with the problems involved in fabricating nano-apertures by FIB milling, describing the ways in which direct milling of a metal film (referred to in this dissertation as direct metal milling or DMM) with a FIB is detrimental to the aperture performance. I then present an improved method of fabrication that protects the metal film by milling through a silicon nitride membrane (referred to as through membrane milling or TMM). I experimentally compare the performance of the TMM aperture to that of the DMM aperture to shown that the transmission is improved in the far-field by nearly an order of magnitude, in close agreement with the simulation predictions. In Chapter 3 I cover an experiment designed to probe the optical near-field of TMM apertures and compares the results to simulation. Chapter 3 is concerned not only with the TMM C-aperture’s performance in relation to conventional round apertures and DMM apertures but also with the absolute intensity enhancement relative to the illuminating light. Having established the high performance of TMM C-apertures in Chapters 2 & 3, in Chapter 4 I apply the improved apertures to imaging and examines their resolving power when used as a source in near-field scanning optical microscopy as well as their non-trivial interactions with the chrome nano-wires being imaged. In Chapter 5 I detail the most complex use of the C-aperture to date. Again using the C-NSOM configuration, the C-aperture is first used to locally heat and record data bits in a Ge$_2$Sb$_2$Te$_5$ phase change layer. Next the same aperture is used to produce a near-field image of the recorded data bits, proving that the C-aperture is capable of subwavelength write/read data storage in phase change media. For verification and to establish the actual size, the recorded data bits are also examined with an AFM and shown to be as small as $\lambda/20$ in width. Finally, in Chapter 6 I conclude with a summary of the major findings and a brief discussion of possible areas of improvement in aperture performance for future work.
2 Aperture Fabrication Using Silicon Nitride Membranes

2.1 Introduction

Nearly all the previous numerical work on nano-apertures has consisted of ideal structures having perfectly sharp edges with 90° corners both inside the aperture and at the entrance and exit surfaces. In contrast, all the experimental work has used FIB milling to fabricate the nano-apertures, which results in an imperfect aperture shape both in the aperture interior and at the surface that was milled with the FIB. This theory-experiment disconnect is a problem for any FIB milled structure where the minimum feature size is comparable to the FWHM spot size of the focused ion beam. These structures are often produced using the simplest FIB milling process where the ion beam is incident directly on the metal surface, a technique we refer to as Direct Metal Milling (DMM). The devices affected by this problem include many plasmonic devices structured on the scale of tens of nanometers and have applications in areas as varied as optical interconnects [61], data storage [62], near-field lithography [63] and biosensing [64].

C-apertures are no exception and although numeric models predict excellent performance the actual fabrication of high quality C-apertures using DMM is nearly impossible, with the device frequently underperforming theoretical predictions by a substantial margin. For
example, in [63] the threshold for photoresist exposed with DMM C-apertures is 2.8 times higher the free space threshold when models predict thresholds one to two orders of magnitude lower [65]. This discrepancy is largely caused by the poor approximation the simulated structure provides to the true shape of DMM apertures. It is well known that the C-aperture geometry and local environment affect the resonant optical transmission [62, 65], but the effects of imperfections due to the realities of FIB milling of plasmonics have been studied only rarely (e.g., [66] examines the effects of tapered side walls).

### 2.2 Sidewall Taper and Edge Rounding

Figure 2.1 diagrams the three primary defects caused by DMM FIB: sidewall taper, surface rounding and interior rounding. The sidewall taper is caused by an angle of incidence variation in the sputter yield that peaks when the angle between the ion beam incidence and the surface normal is 80° but decreases to near zero for angles greater than about 85° [67]. The near zero sputter yield at grazing incidence means that sidewalls remain tapered deep into the milled material. The FIB in the Stanford Nanofabrication Facility (SNF) (that was used in all of the work I present in this dissertation) produces $\theta = 7°$ sidewalls. Surface rounding is caused both by the tendency of sputtering processes to remove sharp edges but also by the ion beam tails eroding material outside the intended mill area. For the C-aperture this is particularly problematic because the ridge receives a double dose from the tails while the two arms are milled. Finally, interior rounding occurs because of the finite size of the focused ion beam (typically ~30 nm FWHM for the SNF FIB). Matteo performed some numerical work into the effects of tapered sidewalls and finite radius of curvature inside the aperture but he treated taper and interior rounding individually and did not examine the rounding that occurs at the mill side surface [59]. He found that independent of other non-idealities, sidewall taper and interior rounding each cause a factor of ~2 decrease in the near-field intensity produced by
the aperture. In this Chapter we will see that, taken together, sidewall taper and the two types of rounding have a much more detrimental effect on aperture performance.

Figure 2.1 – Diagrams of the 3 primary structural non-idealities caused by FIB milling. (a) Ga⁺ beam tails cause surface rounding and angularly dependent sputter yield causes sidewall taper. (b) The Ga⁺ beam profile causes interior corner rounding.

### 2.3 Gallium Contamination

An additional problem with FIB DMM is the contamination of the structure with gallium ions [68]. Noble metals such as gold and silver are popular in plasmonic devices because of their excellent optical properties. When these metals are exposed to a FIB, a thin surface layer becomes impregnated with gallium ions, substantially increasing the optical loss of the metal [69, 70] and dominating when the contamination thickness is on the same scale as the skin depth.

### 2.4 Through Membrane Milling (TMM)

These two difficulties of edge rounding and gallium contamination mean that structures fabricated using DMM lack the sharp corners and low loss metals that are critical to plasmonic devices. To overcome these problems, we propose a method of fabricating apertures that protects the metal layer from damaging gallium ions and beam tails by depositing the metal onto a free standing, silicon nitride membrane. The aperture is milled from the silicon nitride side, through both membrane and metal.
We compare C-apertures of various sizes fabricated with our method of Through the Membrane Milling (TMM) and with conventional DMM using realistic FDTD modeling that includes rounding, gallium contamination and metal surface roughness. The surface roughness plays an important role in output coupling of surface plasmons on the exit surface into the far-field by providing a grating vector for momentum matching. Surface roughness is included in these simulations because our goal is to verify the simulation results using experimental far-field transmission.

The two methods of manufacturing C-apertures are diagramed in Figure 2.2. Figure 2.2a shows the aperture shape used in this study and is the area over which the FIB is raster scanned when fabricating the apertures. In this and subsequent Chapters, the original C-aperture (termed C1 by Shi) has been used rather than the elongated, more optimized designs proposed by Shi [49] or Sendur [50] because of the difficulty in fabricating the later designs with the available FIB tools. Even with the improved method of TMM discussed here, fabricating the best elongated apertures (with ridge widths ~20nm) requires an FIB spot size smaller than is available at SNF. Additionally, experience in fabricating the C1 shape using both DMM and TMM FIB milling has shown that, in order to match the design gap, the milled area between ridge and opposing surface must be reduced in width by ~20%. This is because the ridge tip is exposed to ion beam tails for longer than other parts of the aperture and is thus subject to additional sputtering.

For simulation, the shape in Figure 2.2a is convolved with the Gaussian ion beam current profile of FWHM 33 nm to produce a realistic aperture depth profile. The Gaussian ion beam profile is a good approximation for the most intense portion of the beam although it fails to properly model the wings, which have been shown to have an exponential decay when the intensity reaches $10^3$ times the peak intensity [71]. The aperture is defined by its characteristic size, $\alpha$, which ranges from 40 to 85 nm in this Chapter. Figure 2.2b shows an
example of the simulated gold surface roughness which consists of randomly sized and distributed ellipsoids with a mean diameter of 65 nm and mean height of 4.5 nm. Figure 2.2c shows a cross section (along the dotted line in Figure 2.2a) of a DMM C-aperture with rounded edges and a thin layer of gallium contaminated metal. The substrate layer is likely also contaminated with gallium but this contamination is omitted in these simulations because of the unknown doping levels. The gallium implanted glass may play an important role in blocking light from entering the aperture by direct reflection and by shorting out the aperture modes. Figure 2.2d shows the same cross section for the C-aperture made using the TMM method. As a result of the roughly 10:1 sputter yield ratio between gold and silicon nitride, penetration into the silicon nitride layer is very shallow in the DMM case. Conversely, for TMM, once the FIB breaks through the silicon nitride, the silicon nitride acts as a mask for the gallium beam. This masking allows long mill times that polish the gold sidewalls, reducing the gallium contamination to negligible lateral depths in a manner similar to that used for FIB polishing of TEM samples.

![Figure 2.2 - Schematic drawing of (a) the top view of the mill pattern used in FIB milling and simulation, (b) a sample of the simulated gold surface roughness, (c) a cross section through the tongue of a DMM aperture with AuGa2 lining, and (d) a cross section through the tongue of a TMM aperture. Optical illumination is from below and collection from above](image-url)
in (c) and (d) with measurement of the FDTD calculated electric and magnetic fields at the dotted plane.

## 2.5 Sample Fabrication

The apertures for this study were fabricated in a 75 nm thick silicon nitride membrane onto which a 100 nm thick layer of gold and a 6 nm chrome sticking layer were sputter deposited. The two types of apertures were milled for ~3 s (DMM) and ~30 s (TMM) using a FIB (FEI Strata DB 235) operating at 30 keV and 1 pA beam current.

![Figure 2.3 - SEM images viewed at 52° of tilt of (a) a DMM aperture and (b) a TMM aperture. AFM images of (c) a DMM aperture and (d) a TMM aperture.](image)

The SEM and AFM images in Figure 2.3 show the great improvement in aperture formation with the TMM technique. In Figure 2.3a & c, the tongue of the aperture has been heavily eroded by the tails of the FIB making it appear as a single large hole in the AFM image. The tongue is recessed by about 25 nm from the original gold surface making it nearly useless for applications where the tongue acts as a near-field optical source incident on a plane parallel surface (e.g., optical data storage). Additionally, the extent of the gallium contamination can be seen clearly in Figure 2.3a as the bright area in and around the aperture,
which was shown in [70] to be about 50% Ga by atomic percent. In comparison, Figure 2.3b & d show the TMM aperture is well formed with the metal surface untouched by damaging gallium ions. Although the ridge still suffers from interior rounding, the surface rounding has been completely eliminated (the corner radius is <2.5 nm in AFM images). A careful visual examination of the gold grains in the AFM images in each of the two cases shows that the DMM sample’s gold grain structure has been altered by exposure to the FIB. In Figure 2.3c, none of the very small grains are present within about 100 nm of the aperture perimeter.

2.6 Simulation

The apertures having the same size range as the fabrication (α=40-85 nm) were simulated using an in-house FDTD tool with grid cell sizes of 5 nm in x and y and 1 nm in z. A Drude model fit to tabulated optical constants in a narrow band around 980 nm was used for the simulation of the metals (Au: $\varepsilon_\infty = 11.81$, $\omega_p = 4.19 \times 10^{15}$ s$^{-1}$, $\tau_c = 8.75 \times 10^{15}$ s; Cr: $\varepsilon_\infty = 2.49$, $\omega_p = 5.61 \times 10^{16}$ s$^{-1}$, $\tau_c = 9.49 \times 10^{18}$ s; AuGa$_2$: $\varepsilon_\infty = 14.56$, $\omega_p = 3.15 \times 10^{15}$ s$^{-1}$, $\tau_c = 9.89 \times 10^{-16}$ s [72]). Because optical constants for the Au-Ga composition found in [70] are unavailable, we approximate the contamination as AuGa$_2$. Optical illumination for the simulation consisted of a broad band, plane-wave pulse, x-polarized and centered at 980 nm. A Gaussian FIB current profile having a FWHM of 33 nm (determined from SEM images) was convolved with the mill pattern from Figure 2.2a and used for generating the simulated aperture shape. The simulated mill depth was set by scaling the amplitude of the convolved aperture mill to the metal layer thickness and then multiplying that by the over-mill ratio used in the experiment, which was typically 1.5x. Two DMM aperture cases were simulated, each with the same metal surface shape: one with pure gold throughout the metal layer, and another with a thin layer of gallium contaminated gold (AuGa$_2$) having a maximum penetration depth of 20 nm in the z direction and tapering to zero away from the aperture.
Figure 2.4 – (a) The simulated electric field intensity spectrum 2nm from the metal plane for a DMM aperture with AuGa$_2$ contamination. (b) Same as (a) but for a TMM aperture. (c) The simulated electric field intensity profile 18nm from the metal surface for a DMM aperture with AuGa$_2$ layer and resonant at 980 nm. The overlay in black (white) shows the aperture perimeter at the entrance (exit) of the metal layer. (d) Same as (c) but for a TMM aperture. The entrance and exit are nearly identical and shown in the black overlay. (e) x and y near-field spot FWHM values vs. propagation distance for DMM and TMM apertures. (f) Maximum electric field intensity vs. propagation distance for DMM and TMM apertures.

The electric and magnetic fields were measured 18 nm from the metal exit surface and the simulated near-field intensity is plotted for each milling method in Figure 2.4. Both cases are resonant at $\lambda = 980$ nm as can be seen in Figure 2.4a & b. Figure 2.4c shows a FWHM spot size of 155x92 nm for a DMM C-aperture with $\alpha=40$ nm. Because the DMM aperture lacks the sharp corners at which surface charge can be localized, the optical spot is smeared
out in the x-direction as the surface wave rides along ridge towards the exit, producing a near-field spot substantially larger than the TMM case. Figure 2.4d shows the superior TMM aperture’s FWHM spot size of 66x76 nm for an aperture with $\alpha = 60$ nm. The TMM aperture’s near-field spot is not only 2.2 times smaller on average but is 630 times more intense than the DMM aperture. A comparison between the TMM aperture shown in Figure 2.2d and a fictitious case where the silicon nitride membrane is made contiguous after milling show that the primary effects of extending the aperture channel into the silicon nitride are to shift the resonant $\alpha$ size to larger values (from $\alpha = 50$ nm with the membrane intact to $\alpha = 60$ nm for the structure in Figure 2.2d) and also to increase transmission by about 100%, both due to the lower refractive index and better impedance matching to free space at the aperture entrance. Because the membrane material removed during milling is small compared to the wavelength of light, the effective change in refractive index is a weighted average rather than a simple change from $n = 2$ to $n = 1$.

Figure 2.4e plots the evolution of the near-field spot FWHM as it propagates away from the aperture. The DMM aperture produces a much larger spot at all distances, particularly in the x dimension because of the very large source that constitutes the spot in this dimension. The DMM evolution also lacks the initial high divergence region because the effective emission point for this spot is recessed into the aperture. Note that for the TMM case, the spot size at the aperture exit in the y dimension closely matches the ridge width, $\alpha$, and is an indication that this FWHM can be controlled by sharpening the ridge. Figure 2.4f shows the near-field intensity at $\lambda = 980$ nm and plotted as a function of distance from the aperture exit. The TMM aperture out performs the DMM aperture by a factor of $\sim 10^2$ at all distances which matches the discrepancy found between experimental work (such as in [45]) and simulations of idealized apertures. The difference in Figure 2.4f is also an indication that
the TMM aperture should produce the very high simulated near-field values that have stimulated so much research into these apertures.

### 2.7 Experimental Verification

For verification of the simulation, we measure far-field optical transmission for both fabrication techniques. Far-field transmission is a proxy for near-field intensity [59] and is more accurate when measuring total transmitted intensity since it eliminates distortions due to the presence of a near-field probe tip. The measurement is performed using a confocal configuration. Optical illumination of the sample is from below in Figure 2.2c & d at a wavelength of 980 nm, with a beam polarized along either the x or y-axis and focused at the metal surface to a spot size of 0.9 µm allowing illumination of individual apertures. The apertures are positioned in the x-y plane with a piezo flexure stage and transmission is collected from the gold side with a 0.40 NA microscope objective.

![Figure 2.5 - Aperture far-field transmission scans for apertures ranging in size from α = 40-100 nm. (a) DMM horizontally polarized incident light. (b) TMM horizontally polarized incident light. (c) DMM vertically polarized incident light. (d) TMM vertically polarized incident light.](image)

Figure 2.5 – Aperture far-field transmission scans for apertures ranging in size from α = 40-100 nm. (a) DMM horizontally polarized incident light. (b) TMM horizontally polarized incident light. (c) DMM vertically polarized incident light. (d) TMM vertically polarized incident light.
Figure 2.5 shows the measured far-field transmission for DMM and TMM apertures sourced with light of two polarizations. The horizontally polarized DMM aperture transmission is quite poor with no obvious transmission resonance peak. The peak we expect at $\alpha \approx 40$ nm isn’t visible in this image because transmission from the next higher order mode (with a larger near-field spot) makes a strong contribution for larger apertures. On the other hand, the TMM case has a clear peak at $\alpha = 60-65$ nm, just as predicted from the simulation. Again, the larger apertures with vertical polarization have higher transmission because of transmission through higher order modes. The cross over in far-field intensity occurs at about $\alpha = 70$ nm in this case.

Figure 2.6 - Far-field transmission for the two aperture types normalized to the peak value of the TMM curve. Experimental data are shown as solid curves and simulated data are shown as dashed curves: (a) TMM experimental data, (b) TMM simulated, (c) DMM experimental data, (d) DMM experimental data scaled by a factor of five, (e) DMM simulated without gallium contamination and (f) DMM simulated with gallium contamination. Horizontal error bars show the shift in simulated peak position for a 20% error in sputter yields.

The peak far-field transmission for TMM and DMM apertures are normalized to the TMM peak and plotted in Figure 2.6a, c & d. Figure 2.6a again shows the strong resonance
(\(\alpha=55-60\) nm) of the TMM apertures which is 8.8x more intense than the DMM aperture transmission peak plotted in Figure 2.6c. The DMM data multiplied by five is shown in Figure 2.6d to make the weak resonance peak (\(\alpha=50\) nm) visible. The DMM resonance peak is further suppressed in this experiment relative to previous reports [73] due to the high index silicon nitride backing which provides poor impedance matching to free space at the aperture entrance, shifting the peak position to smaller \(\alpha\) and decreasing peak intensity by about 30%.

The simulated far-field transmissions calculated from the FDTD near-field as described in [74] and normalized to the TMM peak are plotted in the dashed curves of Figure 2.6. Figure 2.6b shows the TMM aperture simulation providing a good match to the experimental data. The small shift in \(\alpha\) peak position likely results from simulated apertures that are slightly smaller than those fabricated due to errors in either the sputter yield or FIB spot size estimates. The DMM simulation with pure gold (Figure 2.6e) has a stronger resonant peak than the experimental data but is still much less intense than the TMM peak. This low intensity is in part due to transmission resonances at the aperture entrance and exit that are very different from one another. Because of the very different aperture cross sections throughout the metal film (see Figure 2.4c) transmitted modes that enter the small aperture entrance must be transformed into a different mode (or tunnel) to emerge at the exit surface, thus making the aperture extremely lossy. Figure 2.6f shows DMM apertures with gallium contaminated gold. The resonance is shifted to \(\alpha=40\) nm, but the transmission is now 6.8x lower than the TMM peak, matching more closely with the experimental values and indicating that gallium contamination does indeed decrease the light transmitted through DMM apertures.
2.8 Summary

In this chapter I have examined the ramifications of the significant disconnect between previous numerical studies and the reality of FIB milled nano-apertures. By simulating the C-aperture shape that can be expected from FIB milling, complete with sidewall taper, and interior/surface rounding, we have shown that many of the exemplary near-field spot properties of C-apertures are lost when the ridge at the aperture’s exits exhibits strong surface rounding. We also find that gallium contamination, clearly present in SEM images of FIB milled apertures, is also detrimental to aperture transmission.

As a solution to the low performance of DMM apertures, we have proposed, simulated, fabricated and tested a method of creating near-IR regime near-field apertures in thin metal films that greatly improves upon conventional FIB techniques. By milling from the silicon nitride side of a gold-on-silicon nitride membrane structure, we protect the gold layer from surface rounding and gallium contamination and are able to form superior apertures. Specifically, we showed via SEM and AFM imaging that the tongue of TMM C-apertures remains sharp. Simulations showed that a TMM C-aperture produces a near-field spot that is simulated as 2.2 times smaller on average and 630 times more intense at the peak than conventional DMM apertures. The improved performance is experimentally verified by measuring the far-field transmission where we found that the TMM transmission is 8.8x higher than the DMM case, in close agreement with the simulated far-field transmission. The preservation of fine features at the metal surface and the protection from gallium contamination that the TMM technique provides is potentially useful in the fabrication of a wide variety of optical near-field structures including bow-tie and fractal apertures, periodic arrays and gratings. The implications for bowtie apertures are particularly good because of the difficulty other researchers have found in fabricating high quality bowties. By milling the center of the gap only lightly, it should be possible to fabricate bowties with sharp tips and
small gaps. In the future, the same technique could be used on sharp-ridged C-apertures, yielding even higher near-field intensities.

In the following chapters, TMM apertures are a critical technological enabler and, as we will see in Chapter 3 where the apertures will be used for high resolution photoresist exposure, DMM apertures often fail to produce a usable near-field spot at all.
3 TPA Probe of the C-Aperture Near-Field

3.1 Introduction to TPA in SU-8 and it’s Applications

Having introduced an improved fabrication technique in Chapter 2 that not only increases the far-field performance of the C-aperture but potentially saves it entirely from the dustbin of impossible-to-fabricate structures, it is important to also test just how well the new aperture performs in the near-field. In Chapter 2 I demonstrated that the far-field performance of the TMM aperture was nearly an order of magnitude better than the DMM aperture and simulations showed that the near-field was 630x more intense. It is now critical to experimentally measure the TMM aperture’s ability to not only produce a small near-field spot (which has already been experimentally demonstrated by several groups), but also to produce an intense near-field spot, capable of affecting a change in an adjacent material.

In order to examine the intensity and shape of the TMM C-aperture’s near-field, we require a suitable, high resolution probe. Often, the size and shape of aperture near-field spots are inspected using other near-field probes, as is the case for previous examinations of the C-aperture’s near-field spot [44]. This is a generally poor method of probing the near-field because the presence of the probe tip can substantially change the intensity and distribution of the very fields it is intended to map. Worse, the resolution of near-field probes, particularly the common round aperture, is rarely better than that predicted for the TMM C-aperture. We are interested in a high resolution mapping, not in reading Braille with a baseball glove.
There are two common alternatives to inspecting a near-field aperture with a near-field aperture: single molecule fluorescence and photoresist exposure. Single molecule fluorescence, where the fluorescing molecules are embedded in a fixed matrix (e.g., spun in a thin layer of PMMA), is often the best option for probing the near-field with high resolution. Unfortunately, in this case it is a difficult experiment because of the general paucity of fluorescent molecules around $\lambda = 800$ nm - 1$\mu$m. This is a challenge because simulations in Chapter 2 showed that the TMM aperture needs to have a characteristic size $\alpha \approx 60$ nm when the exit material is air for $\lambda = 980$ nm. If the exit material is changed to something with a higher index, then the aperture needs to be even smaller. A much more significant issue with this technique is that it requires a scanning setup, where the aperture is translated relative to the layer containing the single molecules. This means that an NSOM-like setup must be used and that each new aperture size to be tested must be separately installed and scanned. This would be a very slow process.

The other common near-field probe is photoresist exposure. The spatial resolution is not single molecule due to molecular diffusion of photoinitiators during resist development but many resists are capable of resolution on the scale of ~10 nm. This technique was used by Wang et al. [45] to compare C-apertures with square and rectangular apertures in a positive resist contact lithography configuration. They found that for an aperture with $\alpha \approx 50$ nm, a 50 x 60 nm hole could be exposed in the photoresist. Unfortunately, the aperture was fabricated using DMM and because they used a UV resist with single photon absorption, they were forced to operate the aperture very far from its main resonance at $\lambda = 355$ nm. As a result (and likely because of the difficulties of contact lithography requiring air gaps of ~10 nm), they exposed only one aperture size and found that, although the exposed hole was $\sim \lambda/7$, it took 2.8 times the optical dose to expose when using the C-aperture compared to direct, far-
field exposure. This exposure threshold result is an order of magnitude from the simulated transmission they reported.

In 2006, Sundaramurthy et al. [40] used a femtosecond Ti:sapphire laser to induce two photon absorption (TPA) in negative photoresist and to examine the near-field of bowtie antennas covered in a thin layer of resist. This technique has the significant advantage of allowing the use of long wavelengths for resonant excitation of the aperture/antenna while still using commonly available, high performing photoresists that require exposure at wavelengths \( \leq 400 \) nm. This process relies on non-linear TPA where an absorption event serves to catalyze the photoresist which is most frequently the epoxy-based negative photoresist SU-8 [75-77]. In addition to the bowtie work done by Sundaramurthy, tip enhanced ANSOM has also been used to expose SU-8 through a TPA process [78].

### 3.2 Setup

In this Chapter, we use the same method of TPA in SU-8 employed by Sundaramurthy et al. and apply it to examining the near-field spot size and intensity of the TMM C-aperture. A diagram of the setup is shown in Figure 3.1. The apertures were TMM with \( \alpha = 45-80 \) nm and fabricated using an 80 nm thick silicon nitride membrane with 100 nm of gold sputter deposited on top. The membrane was spin coated with 75 nm of Micro-Chem SU-8 2002 diluted in cyclopentanone to 2.5 % by volume. The resist layer was prebaked for 5 min at 95 °C. The exposure was accomplished by illuminating the sample from the membrane side with 160 fs pulses at 80 MHz from a Ti:sapphire laser operating at 780 nm. The laser beam was focused to a spot sized 2.3 \( \mu m \) 1/e\(^2\), which was then raster scanned at 25 \( \mu m/s \) over the area to be exposed using a piezo actuated flexure stage (Madcity Labs Nano-Bio2).
Two reference structures were used. First, a membrane with neither gold nor apertures was spin coated with the same SU-8 dilution and used to obtain a free space exposure threshold. Second, a set of round TMM apertures with the same area as the C’s were fabricated on the same membrane structure and also spin coated with the same layer of SU-8 dilution. The apertures were exposed using average beam powers of 0, 0.267, 0.713, 1.160, 1.60, 2.05 and 2.50 mW. The bare membrane was exposed to 20, 35, 40, 50, 75, 100, 150, 200, 250, 300 and 400 mW of average power in order to determine the threshold for photoresist polymerization and damage. After exposure, the samples were given a post exposure bake (PEB) for 1 min at 105 °C and then developed in commercial SU-8 developer for 1 min.

3.3 Simulation of Polymerized Volume

As always, a realistic numerical simulation of the near-fields is important for understanding the experimental results. The setup is modeled as in §2.6 using an in-house FDTD simulation tool except that the exit surface is coated with a 75 nm thick film of SU-8, which is modeled as a non-absorbing dielectric with \( \varepsilon_r = 2.5059 \). Additionally, the z-direction grid size was decreased from the 5 nm cells used in Chapter 2 to 1nm. The surface roughness
is ignored in this model’s because of its small influence on the near-field distribution relative to the C-aperture shape.

Figure 3.2 – (a) Aperture transmission spectrum in the presence of SU-8. (b) Maximum near-field intensity at the laser wavelength of 780 nm. (c) Intensity contour lines along the central x-z slice at the aperture exit. The aperture size is \( \alpha = 45 \) nm. (d) Intensity contour lines along the central y-z slice at the aperture exit. The aperture size is \( \alpha = 45 \) nm.

The results are shown in Figure 3.2. The transmission spectrum (Figure 3.2a) shows that the resonant size is \( \alpha = 40 \) nm, which unfortunately is too small to fabricate consistently. Because we are using two photon absorption for the photoresist exposure the dose scales as the square of the intensity and the incident power required to expose at the ridge of the C-aperture is
where $\eta$ is the near-field enhancement factor plotted in Figure 3.2c & d, $P_{\text{threshold}}$ is the free space power required to expose the resist on the bare membrane, and $P_{\text{inc}}$ is the power incident on the C-aperture. Obviously, this equation reduces to a simple linear ratio and we can state the relationship between incident power and enhancement as

$$P_{\text{inc}} = \frac{P_{\text{threshold}}}{\eta}$$

Equation (3.2) specifies that the aperture in Figure 3.2 should be capable of polymerizing a volume ~20 nm in height and ~50 nm FWHM with an incident power 10 times lower than the free space polymerization threshold. Additionally, the intensity immediately after the aperture reaches a maximum of ~180 times the incident power (Figure 3.2b) indicating that for the $\alpha = 45$ nm aperture, the exposed area should be much larger than for the $\alpha = 80$ nm aperture. For the $\alpha = 80$ nm case, we can expect that the exposure should cover only a small area at the tip of the ridge. Based on Figure 3.2, all the apertures are predicted to produce some polymerization for $P_{\text{threshold}}/P_{\text{inc}} \geq 20$.

### 3.4 Experimental Results

Examining the exposure results for the bare membrane we find that average powers of more than 250 mW either shatter the membrane or damage the photoresist layer via ablation. Powers of 150 and 75 mW are well exposed with a clean perimeter and good adhesion. 50 and 40 mW average powers showed partial exposure with rough edges and some delamination from the membrane surface. Below 40 mW no polymerization was visible. From this we can determine that 40 mW of average power is roughly the free space threshold power for exposure, $P_{\text{threshold}}$. The irradiance for 40 mW is 0.49 MW/cm$^2$ which is smaller but in
acceptable agreement with the threshold value of 2 MW/cm$^2$ reported in [40] given the differences in the incident dielectric layers.

3.4.1 Intensity Enhancement

An AFM was used to examine the size and location of polymerized photoresist adhering to the C-aperture. For average powers 1.60 mW and lower, polymerization was observed only sporadically. At 2.05 mW, polymerization was observed consistently for the smaller aperture sizes but not always for larger apertures indicating that this value is near threshold for the large apertures. 2.5 mW produced consistent polymerization across the whole size range with the height and FWHM of the polymerized volume increasing substantially for smaller $\alpha$ apertures. The threshold value of 2.05 mW average power corresponds to an irradiance of 35 kW/cm$^2$, yielding an experimental demonstration of near-field intensity enhancement over the incident intensity of 20x (and larger for on resonance apertures)!

Figure 3.3a shows an example AFM image of an $\alpha = 45$ nm TMM aperture exposed with 2.5 mW. The polymerized volume is located in the center of the image and partially occludes the aperture as we expect from the simulation iso-surfaces in Figure 3.2c & d. In comparison, Figure 3.3b shows an $\alpha = 80$ nm TMM aperture for the same optical power where the polymerized volume is much smaller and occurs only at the tip of the ridge. In fact, it appears that this exposure only polymerized at the top ridge tip and it is possible that this polymerized volume is a result of the release and diffusion of only a few photoacids. Sundaramurthy et al. found that the minimum feature size was a hemisphere with a 10 nm radius and hypothesized that this was the diffusion distance in 1 min at 105 °C. The height in Figure 3.3b is 11 nm but it is 65 nm FWHM in the x direction. It is possible that the diffusion distance for the membrane sample is larger than that reported in [40] for a glass coverslip.
because of the higher thermal conductivity of the silicon frame on which the silicon nitride membrane rests. However, it is unlikely that the diffusion observed in the x/y versus the z directions is different by a ratio of 7:1. A more likely explanation is that this polymerized volume is the result of the liberation of several photoacids at the top left of the ridge tip but, because ultimately two-photon absorption is a probabilistic process, no photoacids were generated at the bottom left of the ridge tip and we observe polymerized photoresist only at the top left corner of the ridge tip.

Figure 3.3 – (a) AFM image of the polymerized spot produced by an $\alpha = 45$ nm C-aperture. $P_{inc} = 2.5$ mW. (b) AFM image of the polymerized spot from an $\alpha = 80$ nm C-aperture. $P_{inc} = 2.5$ mW. (c) AFM image of a 200 nm diameter circular aperture showing no polymerization. Area is ~3x the area of the $\alpha = 45$ nm C. $P_{inc} = 2.5$ mW. (d) AFM image of an $\alpha = 60$ nm DMM C-aperture with no polymerized spot at the ridge. $P_{inc} = 27$ mW.

For comparison, Figure 3.3c shows a 100 nm radius round aperture exposed to the same 2.5 mW dose and exhibiting no polymerized photoresist. None of the round apertures tested produced a polymerized volume as we should expect from FDTD simulations of equal area round apertures. Also tested were DMM milled C-apertures in a gold film deposited on a
glass substrate. Figure 3.3d shows an example $\alpha = 60$ nm aperture (resonant at $\sim 760$ nm) demonstrating that even at 27 mW of average power, no polymerized photoresist is visible over the ridge.

### 3.4.2 Near-field Optical Confinement

Figure 3.4a plots the polymerized photoresist height values for all the apertures exposed to 2.5 mW and 2.05 mW average powers. The maximum height produced is $\sim 43$ nm demonstrating that the C-aperture is capable of affecting a substantial change in an adjacent medium at least this far into space. The 2.05 mW curve has an average value of roughly 15 nm for larger apertures (those presumed to be just at threshold) meaning that the diffusion distance in this experiment is in fact on average larger than that reported by Sundaramurthy et al. because of the higher thermal conductivity of the silicon frame. Also plotted in Figure 3.4 are the FDTD simulated polymerized heights to which 15 nm has been added to account for the diffusive propagation of photoacids from their generation point. The simulated data values reflect the measured data quite well although they do not reach the same maximum height as the measured values at $\alpha = 45$ nm, likely because of small differences in the actual size of the fabricated apertures due to manufacturing variation (e.g., differing FIB spot size). The perimeter for an $\alpha = 45$ nm aperture should be 594 nm but measurements of the fabricated aperture using AFM images show that the perimeter for the $\alpha = 45$ nm aperture is 534 nm $\pm$ 18 nm which is much closer to the 528 nm perimeter expected for an $\alpha = 40$ nm aperture.

Figure 3.2b shows that with a change in $\alpha$ of just 5 nm around the resonance, the peak intensity can nearly double, substantially increasing the exposed height. On the other hand, for larger apertures $\alpha$ can change substantial with little effect on the exposed height. This strongly suggests that a smaller effective fabricated aperture size is the cause of the discrepancy in exposed photoresist height.
Figure 3.4 – Average polymerized photoresist dimensions as measured by AFM for apertures with $\alpha = 45$-80 nm and dosed with 2.5 mW of average power. The dashed lines show the simulated polymerized volume plus 15 nm (to account for diffusion) for the same optical power and $P_{\text{threshold}} = 40$ mW. (a) Polymerized volume height. Also shown is the average height for 2.05 mW average power. (b) Polymerized x and y FWHM.

Figure 3.4b shows the polymerized volume’s FWHM values in the x and y directions for a dose of 2.5 mW. Here the x FWHM exhibits the same behavior as the height because in the x dimension, the C-aperture approximates a line source and as the intensity increases, more volume falls under the exposure threshold iso-surface. The magnitude of the measured and simulated x-axis FWHM values differ by 20 nm or more. This discrepancy is possibly caused by an amount of surface rounding that is non-negligible compared to the near-field spot size in x. An example of this can be seen in Figure 2.3d where the tip of the ridge is slightly eroded. Any rounding at the ridge tip will expand the spot and lead to a wider spot than predicted by simulation.

The y FWHM is a different story. In this case, we expect the relative extent of the polymerized area to decrease when the near-field intensity decreases, that is, as $\alpha$ increases and the aperture efficiency decreases. However, in the y dimension, the size of the optical spot is set primarily by the width of the ridge such that the effects of decreasing efficiency and increasing ridge width counteract each other leading to a more or less flat FWHM trend versus aperture size. This behavior is apparent in the measured data shown in Figure 3.4b, especially
when compared to the x FWHM behavior. The simulated values show this trend as well, although here the increasing width of the ridge overwhelms the decreasing transmission efficiency in contrast to the measured values where the decreasing efficiency with $\alpha$ dominates. The magnitude of the measured and simulated y FWHM values are a much better match than in the x FWHM case indicating that, in the y direction, the ridge width to exit surface rounding ratio is large and therefore the ridge width is the determining factor in setting spot size.

Finally, we can calculate an estimate of the near-field spot size produced by the TMM C-aperture and compare that to the wavelength of the incident light. Although the $\alpha = 45 \text{ nm}$ aperture is predicted to have the smallest near-field spot, the polymerized volume is overexposed, making it difficult to provide an estimate of the minimum spot size. A better example is the $\alpha = 60 \text{ nm}$ aperture, which produces a small polymerized spot in both x and y of just 65 nm or $\lambda/12$ (see Figure 3.4b). Assuming that the photoacid diffusion distance during the PEB is 15 nm as indicated by the 2.05 mW curve in Figure 3.4a, then the optical spot size is calculated from the measured size as just 35 nm FWHM or $\lambda/22!$ It is important to remember that photoresist exposure is a threshold process and it is possible to create exposures with widths smaller than the nominal FWHM of the incident spot. For instance, by examining Figure 3.2c & d, it can be seen that an incident intensity 100x smaller than the polymerization threshold would expose a volume just 4 nm high, 45 nm in x and 10 nm in y at the ridge tip.

3.5  Summary

In this chapter I have shown that the TMM C-aperture is capable of producing a small near-field spot sufficiently intense to polymerize photoresist. By spin coating C-apertures with a thin layer of the epoxy-based negative photoresist SU-8 and exposing the resist through
the apertures, we polymerized a nano-scale volume, measured its size and compared the result to simulation. Round and DMM apertures were also tested and shown to be incapable of affecting any polymerization. TMM C-apertures were shown to provide a minimum near-field intensity enhancement of 20x over the incident light, matching the simulated values very closely. The shape of the polymerized volume was measured by AFM and the average minimum height of the polymerized hemisphere at threshold was measured as 15 nm indicating that this is the diffusion length during the PEB. The FWHM trends for varying aperture size match the simulated values well, with the x FWHM decreasing as aperture size is increased while the y FWHM remains relatively constant. Finally, the FWHM of the polymerized volume was measured as 65 nm or $\lambda/12$ and if the diffusion distance during PEB is accounted for, the exposed area is roughly 35 nm or $\lambda/22$.

Not only does this experiment demonstrate the ability of TMM apertures to produce a small, intense near-field spot far superior to that produced by round or DMM apertures, but it also proves that FDTD simulations can accurately predict the behavior of C-apertures and it validates the value of C-apertures as a tool for creating near-field optical sources. We have shown that C-apertures are fully capable of producing a near-field spot that is not only (at least) an order of magnitude brighter than the incident light but also that this high intensity is confined to a spot sized about $\lambda/10$. This demonstration is important not only for the following chapters of this dissertation but also for the C-aperture’s application to lithography, optical data storage, nanometer scale inspection and many other tasks requiring an intense, nanometer sized optical spot.
4 Silicon Nitride Membrane NSOM Tips and the C-NSOM

4.1 Introduction

With the continuing reduction in lithographic feature sizes of integrated circuits, it is imperative to be able to identify fabrication and material defects of an equally small size. Historically, production line defect detection/analysis has been achieved using conventional, diffraction limited microscopy but with line widths now < 50 nm this approach is becoming impractical because of the expensive imaging optics, deep UV illumination sources and difficulties associated with imaging in the DUV (e.g., strong UV absorption in air). NSOM is one approach that has been considered as a solution [79] but it is generally considered inadequate due to the very low light throughput intrinsic to deeply subwavelength round apertures. The small amount of light emitted by round apertures in conventional NSOM forces very slow scanning speeds to obtain satisfactory signal-to-noise-ratios and makes the method unacceptable for production line wafer scale defect inspection. The C-aperture is an alternative, which, as we have seen in previous chapters, has two to three orders of magnitude higher brightness than a comparable round aperture while maintaining the very small near-field optical spot required for inspection [65, 73]. We call the configuration where a C-aperture replaces the round aperture in an NSOM setup a C-apertured NSOM (C-NSOM) and
it offers the potential for high speed, high resolution scanning. However, to effectively utilize the C-NSOM it is first crucial to understand how a complicated, resonant nano-aperture such as the C-aperture will perform in the presence of various structures inserted into the near-field, especially those that will be encountered during wafer defect inspection such as nano-scale metallic wires and metallic step edges. In Chapters 2 and 3 I studied the aperture only in the presence of a uniform film, but in this chapter I will examine the response of the aperture to features of finite x-y extent, looking first at chrome nano-wires. The most important distinction between the continuous exit medium and the nano-wire cases is that we expect the near-field of the aperture to interact with, and be strongly affected by, metallic edges.

Conventional NSOM has been studied extensively in the presence of potentially interacting features and, aside from polarization effects [80], changes in local refractive index simply modulate the transmitted light while opaque sample features reduce the transmission. For resonant structures such as the C-aperture, we should not expect the same simplicity. Although the resolution of bowtie apertures and antennas has been explored by measuring the aperture response to point-like features [81, 82], neither the interaction with complicated sample structures nor the use of C-aperture as sources has been studied at all. In this chapter, we show for the first time both the ability to utilize C-apertures in an NSOM configuration and we examine the properties of the C-NSOM for the inspection of chrome nano-wires and step edges. We find that the C-aperture is capable of imaging nano-wires with deeply subwavelength resolution and that transmission is strongly dependent on the orientation and width of the wire.

### 4.2 C-NSOM Configuration

A diagram of the C-aperture used in this chapter is shown in Figure 4.1a. Figure 4.1a also shows the electric field polarization direction relative to the aperture; the polarization-to-
aperture orientation is fixed throughout. The aperture is defined by the characteristic size, $\alpha$, which is set to 50 nm so that the aperture is near resonance for the $\lambda=980$ nm illumination laser as determined by FDTD simulation.

Figure 4.1 – (a) Diagram of the C-aperture dimensions. (b) SEM of a fabricated C-NSOM tip that has been cross sectioned by FIB to show the construction. The corresponding area of (b) is shown in the dashed black line. (c) Diagram of the C-NSOM. Illumination is through the optical fiber at the top. The silicon nitride membrane with truncated pyramidal tip and C-aperture is glued to the cleaved fiber end. The sample to be scanned is positioned under the C-NSOM tip and transmitted light is collected by a microscope objective. The red dotted box shows the volume included in FDTD simulations.

To fabricate the highest quality apertures we again used our technique of through membrane milling (TMM) in which the aperture is milled by FIB from the silicon nitride side of a gold on silicon nitride membrane structure [83]. As was shown in Chapter 2, this milling configuration protects the gold layer from gallium contamination and surface edge rounding, leading to a factor of about 630 increase in near-field spot intensity and a factor of about two reduction in spot size compared to conventionally milled apertures.
4.2.1 Membrane with Truncated Pyramidal Tip

Because the placement of the aperture at the end of an extended tip is critical to avoid collisions between sample surface and extremis points on the probe, we have modified the silicon nitride membrane to include an NSOM suitable tip consisting of a truncated pyramid having dimensions 1.9 µm square at the tip, 2.8 µm tall and 5.8 µm at the base. The design of a probe tip for the C-aperture is further complicated by the fact that the aperture collects light from an area of about \((\lambda/2)^2\). For this reason, in order to maximize the power collection and transmission of the C-aperture at the tip, the C-NSOM tip we have designed has a flat end surface with an area slightly larger than the aperture’s collection area.

The membrane and pyramid fabrication process are shown in Figure 4.2. A double polished <100> silicon wafer was first coated with ~190 nm of low stress silicon nitride using low pressure chemical vapor deposition (LPCVD) (Figure 4.2a). A 5 µm square hole was etched in the silicon nitride by lithographically patterning and then RF plasma dry etching the wafer for 3 min (10 sccm O\(_2\) and 100 sccm SF\(_6\) at 150 mT and 500 W RF power) (Figure 4.2b). The wafer was then immersed in a 40% KOH in water solution at 70 °C to selectively etch the <100> plane, producing a 2.8 µm deep truncated pyramidal hole (Figure 4.2c). The wafer was then covered with an additional 160 nm of silicon nitride, conformally coating the hole and forming the silicon nitride tip (Figure 4.2d). Next, the silicon nitride on the opposite side of the wafer was patterned and then RF plasma etched with the same parameters used for the 5 µm hole (Figure 4.2a), this time to produce a 650 µm square hole in the silicon nitride coating (Figure 4.2e). Finally, the wafer was immersed in a 30% KOH in water solution at 85 °C for 7 hours to etch through the entire wafer and release the ~140x140 µm membrane with the truncated pyramidal tip at the center (Figure 4.2f).

After evaporating 120 nm of gold onto the tip protrusion side of the membrane (bottom surface in Figure 4.2f), the C-aperture was formed by milling from the silicon nitride
side of the membrane using an FEI Strata DB 235 focused ion beam for 27.4 s (operating current of 1 pA, beam spot FWHM of ~33 nm). Note that the TMM fabrication used here is also applicable to AFM type cantilevers, offering the potential for massively parallel imaging [84].

Using UV cure adhesive (Norland Optical Adhesive 81), the membrane was glued to a cleaved optical fiber (Corning HI-980; mode field diameter of 2.1 µm), then broken so that the membrane free of its supporting silicon frame to form the C-NSOM fiber probe tip (Figure 4.1b & c). Illumination was through the fiber and collection was with a microscope objective (0.25 NA). The tip-to-sample gap was held constant at a 2-10 nm separation using tuning fork based shear-force feedback [85]. Positioning in the x-y plane was with a piezo flexure stage with 0.2 nm resolution (Madcity Labs Nano-Bio 2) and z positioning was with a 6.1 µm travel piezo stack. The laser source was a voltage pulsed telecom DFB (λ = 980 nm) with polarization adjusted to be x-polarized at the aperture. C-NSOM scans used 0.6 µW of DC optical power and a constant scan speed of 1 µm/s. The chrome nano-wires were fabricated on a glass microscope cover-slip using a standard electron-beam lithography lift-off process. The wires were evaporated to 20 nm thick and were 3 µm long having widths of 50, 85, 130 and 235 nm. Inter-wire spacing was 3 µm.

![Figure 4.2 – Fabrication of a silicon nitride membrane with truncated pyramid.](image)

(a) A double polished <100> silicon wafer is LPCVD coated with 190 nm of low stress silicon nitride. (b) A 5 µm square hole is lithographically patterned and etched in the top surface using an O₂ and SF₆ RF plasma etch. (c) The wafer is immersed in 40% KOH in water at 70 °C to produce a pyramidal mold. (d) The wafer is coated with an additional 160 nm of LPCVD low stress nitride to define the pyramidal tip in silicon nitride. (e) A 650 µm square hole is...
etched in the backside silicon nitride as in (b). (f) The silicon is etched by immersion in 30% KOH in water at 85 °C to release the membrane.

4.3 Chrome Nano-Wires

4.3.1 Simulation

To more fully understand the system, we performed FDTD simulations of the aperture and nano-wire system and propagated the near-field response into the far-field using the method described in Ref. [86]. By running multiple simulations with varying aperture-to-wire x-y displacements and recording the far-field transmission for each displacement we simulated C-NSOM line scans across the width of the nano-wire for comparison to experiment. The area included in the simulation is shown by the red dashed box of Figure 4.1c. The simulation parameters were a grid size of 5 nm in x and y, and 1 nm in z. Aperture-to-wire separation was 2 nm in the z direction and the optical illumination was a plane wave pulse. A Drude model was used for simulation of metals with the parameters derived from a fit to tabulated values in Ref. [87] (Au: \(\varepsilon_\infty = 13.0590, \ \omega_p = 1.4558 \times 10^{16} \ \text{s}^{-1}, \ \tau_c = 9.0261 \times 10^{-15} \ \text{s}; \ \text{Cr}: \ \varepsilon_\infty = 2.4901, \ \omega_p = 8.8635 \times 10^{16} \ \text{s}^{-1}, \ \tau_c = 9.4751 \times 10^{-18} \ \text{s})]. The results of this virtual NSOM (v-NSOM) simulation are plotted in Figure 4.4 for comparison against the experimental values.

4.3.2 Experimental Results

The C-NSOM images were constructed by bringing the truncated pyramidal tip close to the chrome nano-wire while maintaining the distance using the shear-force feedback control loop and raster scanning over the area of interest. During the scan, the transmitted light and the z-axis feedback loop position were measured at each point to build simultaneous AFM and optical transmission images. Figure 4.3a shows an example C-NSOM scan of wires oriented parallel to the electric field polarization (hereafter transverse magnetic – TM; aperture orientation in inset). The wires are clearly visible with transmission reductions of \(\sim 10\%\) for
the 50 nm line and ~30% for the 235 nm line. The box seen around each wire in the optical image is an artifact caused when the flat end of the pyramidal tip leaves the glass substrate surface and lifts up over the nano-wire, changing the local refractive index and consequently the transmission. The lifting can be seen in the AFM images shown in Figure 4.3b & d. Also visible in the lower portion of Figure 4.3d are AFM images of the tip shape, complete with the C-aperture imprint, caused by scanning over a sharp piece of debris that has become stuck to the sample surface. When this happens, the debris acts as a spontaneous AFM tip, facilitating the formation of the image of the C-NSOM tip.

Figure 4.3 – (a) C-NSOM scan of chrome nano-wires oriented parallel to the polarization. (b) The corresponding AFM image for (a). (c) C-NSOM scan of chrome nano-wires oriented perpendicular to the polarization. Debris obstructs the center of the 130nm nano-wire; the right hand side is unobstructed. (d) The corresponding AFM image for (c).

In the TM orientation, the wires produce symmetric dips in transmission and the cross sections along each of the wires closely match the simulated values (shown in Figure 4.4a). Simulations show that the amplitude of peaks and troughs is strongly affected by small changes in the aperture-to-wire separation and the observed difference in transmission between experiment and simulation is likely due primarily to a slightly larger separation in the
experiment than in the simulation. The measured FWHM values shown in Table 4.1 are roughly equal to the wire width plus a 100 nm offset, indicating that the resolution is ~100 nm ($\lambda/9.8$) in the y-direction.

In contrast to the dips seen in Figure 4.3a, the example scan in Figure 4.3c shows nano-wires oriented perpendicular to the electric field polarization (transverse electric – TE). Here, the wider (130 and 235 nm) wires also produce dips, but the 50 and 85 nm wires exhibit a strong interaction with the C-aperture and, for particular displacements, produce increased transmission. This interaction can be seen more clearly in Figure 4.4c showing line-scans across the wires. The scan over the 50 nm wire has the strongest increased transmission, which manifests as a peak to the left of the dip. As the wire width increases, the peak becomes weaker and is eventually subsumed by the dip, leaving a cusp to the left of the dip. For the 50 nm wire, the measured width is 124 nm (Table 4.1), and taking the resolution as the measured width minus the wire width, the resolution is 74 nm ($\lambda/13$) in the x-direction.

<table>
<thead>
<tr>
<th>Wire width (nm)</th>
<th>TM - Measured</th>
<th>TM - Simulated</th>
<th>TE - Measured</th>
<th>TE - Simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>146</td>
<td>125</td>
<td>124*</td>
<td>100*</td>
</tr>
<tr>
<td>85</td>
<td>194</td>
<td>157</td>
<td>88</td>
<td>90</td>
</tr>
<tr>
<td>130</td>
<td>228</td>
<td>207</td>
<td>159</td>
<td>151</td>
</tr>
<tr>
<td>235</td>
<td>329</td>
<td>305</td>
<td>231</td>
<td>258</td>
</tr>
</tbody>
</table>

Table 4.1 – Measured and simulated transmission dip FWHM for various nano-wire widths in the TM and TE orientation. *For the 50 nm wire in the TE orientation, width is the distance between the right hand half minimum of the dip and the left hand half maximum of the peak.

4.3.3 Aperture-to-Wire Coupling

The simulated line scans shown in Figure 4.4a & b match the shape of the experimental data well aside from the amplitude and by looking closely at the data available from the simulation we can understand the origins of these anomalous transmission peaks. The diagrams in Figure 4.4c-e show free charge displacement at various aperture-to-wire...
displacements. FDTD simulations of the charge density are shown in Figure 4.4f-h. As we know from Chapter 1, during normal operation, an electric dipole is established between the aperture ridge and the opposite interior wall (Figure 4.4c,f) and the high charge density at the end of the ridge sources emission into the far-field. For TM wires, the transmission is reduced because the nano-wire behaves as a mirror and obstructs emission.

For TE wires with widths roughly matching the aperture gap width, the aperture dipole induces another dipole across the width of the wire that then re-emits (Figure 4.4d,g). When the wire is under the aperture ridge (Figure 4.4e,h) transverse coupling does not occur, and we observe a transmission dip due to obstructed emission. For wider TE wires, the aperture dipole is never efficiently coupled to the wire, but exhibits a cusp where the right hand wire edge encounters the aperture ridge and partially couples.

Figure 4.4 – (a) C-NSOM line scans across TM chrome nano-wires. Data is offset vertically by 0.15, 0.1, 0.05 and 0.0 for the 50, 85, 130 and 235 nm wires respectively. (b) C-NSOM line scans across TE chrome nano-wires. Data offset vertically as in (a). Arrows indicate the positions of (c-h). (c,f) Diagram (c) and FDTD simulation of normalized charge density (f) of a cross section along the red dashed line of Fig. 1(a) showing the charge displacement in
the C-aperture during normal operation. Illumination is from the top. Emission to the far-field is shown as red arrows. (d,g) Diagram and simulation of operation with a chrome wire bridging the gap. (e,h) Diagram and simulation of operation with a chrome wire offset away from the gap.

4.4 Chrome Step Edge

4.4.1 Fabrication

Another related structure of interest is the metallic step edge. To measure the response of C-apertures to the step edge we fabricated a sample by depositing polystyrene spheres in water on a glass coverslip (see Figure 4.5). When the water evaporates, the spheres form large aggregates that can be used as randomly shaped liftoff masks by evaporating metal over the top and removing the spheres (and the metal on top of them) by sonicating the sample in organic solvents.

Figure 4.5 – Fabrication of chrome step edge. (a) Evaporation of water leaves an aggregate of polystyrene spheres. (b) Evaporation of chrome. (c) Removal of the polystyrene spheres and any chrome on top. (d) The liftoff leaves protrusions along some of the edges.

If the thickness of the evaporated metal is too large, the edges can exhibit point protrusions where the metal on top of the sphere forms a connection with the metal on the surface. During the liftoff process, these connections tear leaving protruding metal as shown in Figure 4.5d. Normally this is an undesirable property as it causes rough edges; however, if
the protrusions are small and few in number, they can be used to examine the aperture’s resolution as a function of tip-to-sample distance.

### 4.4.2 Transmission Profile

The behavior of the aperture as it scans over a step edge is expected to be different from the TE nano-wire case because the step affords no opportunity for strong coupling to a dipole in either the x or y direction and is thus mostly non-interacting (or TM-like). Instead, the transmission profile resulting from scanning the step edge can be approximated as the intensity profile of the probe tip convolved with the step. This understanding is only an approximation (although Figure 4.6c shows that it is a good one) because although the aperture doesn’t strongly couple to the edge, there are still changes in the electric field pattern near the edge such that the aperture intensity profile is not constant with position as is supposed by a simple convolution.

![Figure 4.6](image)

**Figure 4.6** – **(a)** The intensity profile of the C-NSOM probe tip. Transmission of the illumination spot through the gold layer produces a large background envelope (mode field diameter of 2.1 µm). The small near-field spot produced by the C-aperture is superimposed on the wider leakage background. **(b)** The transmission profile that results from scanning over a step edge; essentially a convolution of the curve in (a) with a step. **(c)** Experimental observation of the center portion of the (b) with a fit to the data using Equation (4.1).

Figure 4.6a shows the intensity profile produced in the near-field of the truncated pyramidal probe tip. There is a large and wide Gaussian background caused by leakage of the
illumination beam through the gold layer. The size of this spot is 2.1 μm, the same as the mode field diameter of the Corning HI-980 fiber used to illuminate the apertures. The magnitude of the background envelope is set by the transmission coefficient through the 160 nm thick silicon nitride membrane with 120 nm of evaporated gold and is \( \sim 2.5 \times 10^{-3} \). In addition to the wide Gaussian, there is the smaller peak produced by the near-field spot of the C-aperture. The magnitude and width of the C-aperture peak are determined by the properties of the C and, in terms of the intensity compared to the background when observed in the far-field, how effectively the near-field spot is coupled into the far-field.

Figure 4.6b shows the transmitted intensity profile as the aperture is scanned over a step edge. The profile consists of a large Gaussian derived cumulative distribution function from the leakage background superimposed on a smaller one. In order to fit this curve and extract an effective width, we use a linear approximation to the center portion of the larger distribution and a sigmoid function for the smaller distribution. The resulting fit equation is

\[
y(x) = y_0 + mx + \frac{a}{1 + e^{-k(x-x_c)}}
\]

where \( x \) is position, \( y_0 \) is a vertical intensity offset, \( m \) is the slope of the large envelope linear approximation, \( k \) is the width parameter of the sigmoid and \( x_c \) is the center position of the sigmoid. Figure 4.6c shows example data of a C-NSOM scan over a chrome step edge and a linear plus sigmoid fit to that data showing excellent agreement between model and data.

The wide background sigmoid and the narrow one caused by the C-aperture can be seen clearly in the C-NSOM image shown in Figure 4.7a. The shape of the transition is particularly obvious along the step edge at \( x = 4 \mu m \) where there is chrome on the left hand side and bare glass to the right. The simultaneously acquired AFM image is shown in Figure 4.7b. The many edge protrusions of varying height along the chrome perimeter are manifested as bright square replicas of the probe tip in the AFM image. The shape of the C-aperture can
also be seen in the center of the squares, indicating that the protrusions are point-like. By correlating the raised squares in the AFM image to the optical image, we see that the greater the protrusion height, the broader the transition width. Good examples of this effect are at \((x,y) \approx (4,8) \, \mu m\) and \((9,10) \, \mu m\) where the protrusions are particularly large and the optical transition width is visibly increased compared to adjacent line scans.

Figure 4.7 – (a) C-NSOM image of chrome step edges. (b) AFM image of the same structure as (b). Note the varying height of individual protrusions along the chrome edge.

### 4.4.3 Near-Field Spot Size vs. z Propagation

The chrome islands in Figure 4.7 are particularly rich in x-axis edges so we choose to examine the C-aperture resolution only in the x direction as a function of tip-to-sample distance. Choosing several transitions along the chrome island’s perimeter so as to span a wide range of tip-to-sample heights, we first measure the tip-to-chrome distance from the AFM image and then fit the optical transition data using Equation (4.1). The result is plotted in Figure 4.8 where the width is measured as the 10%-90% width for the sigmoid component of the fit. The increase in sigmoid width with increasing distance is clear. A linear fit to the data shows that the width of the near-field optical spot increases rapidly, with a slope of 2.9 nm per nm of z propagation. The \(x = 0\) intercept yields a minimum resolution of the aperture near a chrome step edge of 101 nm. The 10%-90% sigmoid width corresponds to a near-field spot FWHM of 10% smaller than the sigmoid width if a simple convolution is assumed. This
implies that the minimum interaction distance of the near-field spot with the chrome step edge is 91 nm FWHM or \( \lambda/11 \).

![Figure 4.8 – Sigmoid widths for fit to C-NSOM line scans over chrome edges with protrusions of varying heights (black). Linear fit to the sigmoid widths (red). Simulated FWHM along the x axis plus a 101 nm offset (blue).](image)

For comparison to the slope of the transition width data values vs. tip-chrome distance, the FDTD simulated x FWHM values as a function of distance for a C-aperture sourcing into free space are also plotted in Figure 4.8. Because the simulation is idealized with an infinitesimally sharp ridge corner at the exit, the x FWHM is essentially zero at the surface and the value plotted in Figure 4.8 has been increased by 101 nm to facilitate comparison with the data. The 101 nm minimum transition width is likely the result of one or more of the following: (1) actual tip-to-sample separation larger than zero, (2) exit surface rounding at the ridge, or (3) an increase in spot size caused by the interaction between step edge and aperture. (1) is probably a small effect (~10 nm) given that the AFM image shows no protrusions from the C-NSOM probe itself and there are several areas with step edges without protrusions. (2) may have an effect although it seems unlikely that a 5-10 nm radius corner would produce a 100 nm increase in spot size. (3) seems the most likely reason for the discrepancy given the strong distortion observed in simulation when a metal edge is brought
near to the aperture. Under these conditions, the approach of the step edge has a larger effect on the aperture transmission than would otherwise be suggested by a linear interaction with the near-field intensity.

Looking only at the slopes, we see that the data and simulation are an excellent match, demonstrating that the simulation accurately predicts the expansion rate of the optical spot at increasing distance from the metal surface. This data is also a clear warning about the importance of approaching the aperture as close as possible to the sample surface to achieve small optical spots.

### 4.5 Summary

We have demonstrated for the first time optical imaging using a C-aperture as a near-field probe. We imaged 50 nm ($\lambda/20$) features with resolution of 74 nm in x and 100 nm in y. TM chrome nano-wires appear semi-opaque, producing a dip in far-field transmission. TE nano-wires having widths roughly equal to the aperture gap produce an increased transmission peak to the left of the dip due to electric dipole coupling between the aperture and nano-wire. TE wires having widths larger than the gap do not efficiently couple to the aperture dipole but do produce transmission dips with a cusp to the left of the dip that is caused by partial coupling.

Finally, we imaged chrome step edges using the C-NSOM and demonstrated that the C-aperture is capable of imaging with < 100 nm resolution ($\lambda/9.8$) the position of metallic edges. Further, randomly distributed point protrusions along the edge of the chrome step edge were used to measure the C-aperture’s resolving power at various tip-to-sample separations and this data was used to determine that the expansion rate of the near-field optical spot is 2.9 nm per nm of z propagation which is in close agreement with simulation. This rate of
expansion is critically important for understanding the tolerances on tip-to-sample separation and underlines the necessity of minimal separation for producing small near-field optical spots.
5 C-Aperture Based Optical Data Storage

5.1 Introduction

So far, I have described, simulated and experimentally demonstrated many of the properties of the C-aperture. I’ve shown that the C-aperture exhibits resonant transmission, and is capable of producing a small and intense near-field spot. I’ve also shown that subwavelength structures can be imaged with the C-aperture and that the interactions of the aperture with the subwavelength structure are eminently predictable through FDTD simulation. We can now apply all of these lessons to the problem of optical data storage in a demonstration of the ability of the C-aperture to both record and read deeply subwavelength data bits.

Lens-based optical data storage has enjoyed incredible success from the development of compact disks to the current generation of Blu-ray disks. Unfortunately, as described in Chapter 1, the data density of these lens-based methods is fundamentally restricted by the diffraction limit, with optical spot sizes and data bit sizes of $\sim 0.5\lambda/NA$, where $\lambda$ is the wavelength of incident light and $NA$ is the numerical aperture of the lens. As seen in the progression from CD-ROM ($\lambda=780$ nm, $NA=0.50$) to Blu-ray ($\lambda=405$ nm, $NA=0.85$), the historical approach to increasing data density has been to decrease $\lambda$ and increase $NA$, but these improvements have reached practical limits. Specifically, $\lambda$ is limited by materials
available for inexpensive, long-lived semiconductor lasers, and $NA$ is limited to 1.0 outside the near-field.

A variety of solutions to this problem have been explored using near-field optical effects, several of which have already been discussed. Common approaches include Solid Immersion Lenses (SILs) [88-90], super-Resolution Near-field Structures (super-RENS) [91, 92], sharp plasmonic tips [93, 94] and apertures [29, 50, 55, 95, 96]. Among these, apertures offer particular promise. In particular, the resolution is not limited by the refractive index of available materials (a limitation of SILs), there are few extremely challenging and poorly understood materials engineering problems to be solved (a limitation of super-RENS), and there is no large background signal to degrade readout SNR (a limitation for sharp plasmonic tips).

This isn’t to say that the apertured path is free of obstacles; if it were, apertured recording would have been commercialized years ago. The history of aperture based recording research and a careful look at the previous results is instructive as to what the problems are. At least one claim of sub-100 nm bit size recording in a phase change medium using round apertures has been published [97]. Unfortunately, since publication in 1996, this result has not been repeated or improved upon. Looking closely at their results, it can be seen that the areas where they claim to have written bits is increased in thickness by 50 nm (compared to a recording layer that is just 30 nm thick). Bubbles that bulge the surface can sometimes be formed during recording but typically take many ($>10^3$) record/erase events to appear and this kind of surface deformation is a clear indication of collision or other surface damage rather than data bit recording and strongly connotes a false positive. Moreover, it was later shown that heating of the recording medium was dominated by optical absorption in the aluminum tip with subsequent conductive transfer to the medium rather than by direct optical absorption heating by the near-field optical spot [98]. This last result illustrates the
problem with conventional aperture-based recording, namely that low power throughput makes high density optical recording nearly impossible. To solve this problem, we change the aperture shape from the conventional circle to a C-shape, boosting power transmission by up to three orders of magnitude while maintaining a small near-field optical spot [49, 99] and bringing aperture-based optical data storage within reach using commercially available low-cost mass produced lasers enabling fast recording times commensurate with the high data density.

5.2 Experimental Setup

To study high density data storage with C-apertures, we used the same near-field scanning optical microscope (NSOM) with a specially designed C-aperture fiber probe tip (C-NSOM) as introduced in Chapter 4. Aperture dimensions are shown in Figure 5.1a. Again adapting the previously described technique from Chapter 2 for fabricating high quality C-apertures by milling through a silicon nitride membrane with a FIB [100], we use the same silicon nitride membrane that includes a truncated pyramidal tip protruding from the center (Figure 5.1c and d). The pyramidal tip had dimensions 1.9 µm square at the tip, 2.8 µm tall, 5.8 µm at the base, 160 nm thick and was sputter coated with 120 nm of gold (see Figure 4.2 for details). The apertures were milled for 27.4 s using an FEI Strata DB 235 FIB operating at 1 pA. The membrane was then glued (Norland Optical Adhesive 81) to an optical fiber (Corning HI-980) forming our C-NSOM probe tip. The gap between fiber probe tip and recording medium was maintained using conventional quartz crystal tuning fork based shear-force feedback [101]. Positioning in the x-y plane was achieved with a piezo flexure stage with 0.2 nm resolution (Madcity Labs Nano-Bio 2) and z positioning was with a 6.1 µm travel piezo stack. The laser source was a voltage pulsed telecom DFB ($\lambda = 980$ nm) with polarization adjusted to be x-polarized at the aperture. C-NSOM scans used 0.6 µW of DC
optical power and a constant scan speed of 1 µm/s. Transmitted light was collected with a 0.25 NA microscope objective and measured with a NewFocus 2151 Femtowatt detector. For the recording medium, we used the chalcogenide compound Ge$_2$Sb$_2$Te$_5$, which is commonly found in DVD-RW media [102]. Data bits are stored by changing the phase from amorphous to crystalline by heating the Ge$_2$Sb$_2$Te$_5$ above the glass transition temperature ($T_g \sim 150^\circ C$), resulting in large reflectivity and transmission changes. We used a 20 nm thick amorphous (a-) Ge$_2$Sb$_2$Te$_5$ layer that was deposited by RF sputtering in a thin film stack detailed in Figure 5.1d.

Figure 5.1– C-NSOM setup. (a) The aperture shape used in this study, defined by the characteristic size, $\alpha$. (b) FDTD simulated near-field spot cross sections normalized to $|E_{inc}|^2$ in the center of the a-Ge$_2$Sb$_2$Te$_5$ layer. (c) An SEM image of an FIB cross sectioned pyramidal tip corresponding to the dotted box area in (d). Tilt angle is 52°. (d) The assembled C-NSOM probe tip shown oriented above the recording medium (not to scale). Light is delivered from the top. The recording medium consists of (from top to bottom): a silicon nitride capping layer to prevent Ge$_2$Sb$_2$Te$_5$ oxidation, the a-Ge$_2$Sb$_2$Te$_5$ recording layer (shown with c-Ge$_2$Sb$_2$Te$_5$ recorded bits), a silicon nitride thermal buffer layer and a titanium heat sink layer.
5.3 Simulation of C-Aperture Near GST

Because of the close proximity of the recording media, the aperture characteristic size, \( \alpha \), must be scaled such that the aperture is resonant near the laser wavelength of 980 nm, maximizing the energy delivered [50]. We found that \( \alpha = 50 \) nm is near resonance by simulating the behavior of the electromagnetic fields using custom Finite-Difference Time-Domain (FDTD) [86] software that allows modeling of complicated FIB fabricated shapes (simulation volume is shown in the black dashed box of Figure 5.1d). Figure 5.1b shows the normalized electric field intensity profile in the Ge\(_2\)Sb\(_2\)Te\(_5\) layer. The intensity profile along x shows significant kurtosis, presaging that overexposing will result in data bits that are elongated in the x-direction while near-threshold recorded bits will be symmetric. The simulation illumination was a plane wave, broad band pulse, polarized along the x-axis. The simulation grid was 5 nm in x and y and 1 nm in z with perfectly matched layer (PML) boundary conditions on all sides. The metals were represented using a Drude model (Au: \( \varepsilon_{\infty} = 13.0590, \omega_p = 1.4558 \times 10^{16} \) s\(^{-1}\), \( \tau_c = 9.0261 \times 10^{15} \) s; Ti: \( \varepsilon_{\infty} = 1.00, \omega_p = 7.4181 \times 10^{16} \) s\(^{-1}\), \( \tau_c = 1.0931 \times 10^{-17} \) s). The Ge\(_2\)Sb\(_2\)Te\(_5\) was simulated as a lossy dielectric with \( \varepsilon_{\text{amorphous}} = 20.98 + i\cdot7.72 \) and \( \varepsilon_{\text{crystalline}} = 5.047 + i\cdot37.11 \). The simulated aperture shape was created by convolving the aperture pattern in Figure 5.1a with a Gaussian of FWHM 33 nm (matching the experimental FIB instrument value). Far-field transmission was calculated as described in reference [86]. To simulate C-NSOM response curves, a crystalline bit (c-Ge\(_2\)Sb\(_2\)Te\(_5\)) with an elliptical cross section in x-y and straight sidewalls in the z direction was inserted into the a-Ge\(_2\)Sb\(_2\)Te\(_5\) layer. The far-field transmission from multiple simulations was recorded as the bit was translated in the x-y plane relative to the aperture. This approach is identical to that used in Chapter 4 to simulate the C-NSOM scans over the chrome nano-wires.
5.4 Experimental Results

5.4.1 Optical Pulse Length

We recorded data bits using optical power in the fiber from 13.0 mW to 21.6 mW and pulse lengths, $t_{\text{pulse}}$, ranging from 1 µs to 1 ms. The time required for phase change in Ge$_2$Sb$_2$Te$_5$ is <50 ns [103]; the 1 µs minimum $t_{\text{pulse}}$ is a limitation of our laser switching time constant, not a physical limitation on data rate. This can be seen experimentally by noticing that although the pulse lengths vary by 3 orders of magnitude, the power required for writing only varies by a factor of 2! This is a consequence of the thermal propagation constant of the recording stack being quite short and, for long pulses, results in substantial conductive loss of heat to surrounding areas during the lifetime of the optical pulse. For shorter pulses, there is insufficient time for the heat to spread and lower optical powers (and in fact lower total energies) are required to change the Ge$_2$Sb$_2$Te$_5$ phase. This effect can be simulated using a finite element solver. Figure 5.2 shows the heating at the center of the data bit for two different pulse lengths (solved using the commercially available COMSOL Multiphysics FEM tool). The 10 µs pulse is shown in red and has the same parameters as used in the experiment. The black curve is a 100 ns pulse where the $1/e^2$ width far-field optical spot used to illuminate the C-aperture has been reduced from 2.1 µm to 1.0 µm to represent a spot typically produced by a commercial DVD lens. Note that in going from a 10 µs pulse, to a 100 ns pulse with a slightly smaller illuminating width, the same heating can be achieved with just 12.5 mW of input power compared to the 19.3 mW needed in the experiment for the longer pulse. This result is an indication that, although the experiment was limited to relatively long pulse times by the available equipment, the C-aperture and the recording media are fully capable of operating with short pulses at low optical power levels that are commensurate with current optical storage systems.
Figure 5.2 – Finite Element Simulation of Temperature. The temperature in the center of the Ge$_2$Sb$_2$Te$_5$ layer simulated using COMSOL Multiphysics for pulse lengths of 100 ns and 10 µs with far-field illumination spot sizes ($1/e^2$) of 1.0 µm and 2.1 µm respectively. The 10 µs pulse power is set at the threshold power and illumination spot size from experiment. For the 100 ns pulse, the far-field illumination spot size has been reduced to a value attainable with an NA = 0.6 lens (the same as used in DVD) and the power set to produce the same temperature as the 10 µs pulse.

5.4.2 Optical Readout of Recorded Bits

After recording, we read the local optical transmission by C-NSOM. An example of the optical response is shown in Figure 5.3a. Cross sectional curves are shown in Figure 5.3b and c with FWHM of 195.8x131.7 nm. FDTD simulated C-NSOM cross sections in Figure 5.3b and c are for an elliptical crystalline bit with size 60x50 nm in x-y. Figure 5.4 shows the simultaneously acquired AFM data for the area scanned in Figure 5.3a. The surface is flat to within ~4 nm and shows clearly that the reduced transmission spots in Figure 5.3a are not due to z-motion artifacts caused by surface deformation [104]. If the C-aperture is to be used in a spinning disk/flying head configuration, then the flat surface after writing is crucial to avoid collisions between the disk and the recording head/aperture.
Figure 5.3 – Optical and AFM bit readout. (a) A transmission C-NSOM image of bits recorded with 10 µs, 19.34 mW pulses. (b,c) Cross sections along the white dashed line in (a). FDTD simulated C-NSOM cross sections for an elliptical crystalline bit with x-y dimensions of 60x50 nm are shown in as circles. A halo of increased transmission can be seen around the transmission dip in (a-c).

Figure 5.4 – Surface topography after recording. AFM data of the capping layer that was simultaneously acquired with the C-NSOM scan shown in Figure 5.3a. The surface is flat to within ~4 nm, demonstrating that the transmission changes measured in Figure 5.3a are optical in origin. Additionally, because future near-field recording systems must incorporate read/write flying heads similar to those found in today’s hard drives and operating fly heights on the order of 10 nanometers, it is extremely important that there is no surface deformation after writing to prevent head crashes.
5.4.3 Optical Readout Halo

The simulation shows excellent agreement with experimental data and aids in understanding the faint halo of increased transmission around the optical response dip (Figure 5.3a-c). This halo also occurs in simulations, indicating that it is not caused by artifacts such as surface deformation [104] or ‘priming’ of the amorphous state [105]. Instead, the halo is generated by surface plasmon (SP) waves that propagate radially from the aperture along the exit surface of the metal film [106, 107]. In Reference [106], the observed SP standing wave was caused by interference between the SP wave propagating along the exit surface and the partially transmitted far-field illumination light, resulting in a standing wave with wavelength $\lambda_{sp}$. In contrast, the data Figure 5.3 has the halo peak at roughly $\lambda_{sp}/2$ suggesting a different process is responsible: when the SP wave encounters the crystalline bit, it is partially reflected back to the aperture forming a cavity that constructively interferes when the cavity length is equal to half the SP wavelength. The field inside the cavity is scattered into the far-field by the aperture edges and the recorded bit, both of which act like apertureless NSOMs [15]. The SP wavelength is determined by a restatement of the dispersion relation in Equation (1.12).

$$\lambda_{sp} = \frac{\lambda}{\sqrt{\varepsilon'_m + \varepsilon_{d,eff}}}$$  \hspace{1cm} (5.1)

where $\varepsilon_{d,eff}$ is the effective dielectric constant and is found by calculating a weighted average of the refractive indices, $n(z)$, adjacent to the metal layer and through which the SP electric field propagates. The weighting factor is the electric field intensity, $|E|^2$ and the effective refractive index is then [108]

$$n_{eff} = \frac{\int_0^\infty n(z)|E(z)|^2dz}{\int_0^\infty |E(z)|^2dz}$$  \hspace{1cm} (5.2)

Because $|E|^2$ is known from FDTD simulations, the calculation is straightforward and for $\lambda = 980$ nm, $n_{eff} = 1.066$ and $\varepsilon_{d,eff} = n_{eff}^2 = 1.138$. Plugging into Equation (5.1) yields
\( \lambda_{sp}/2 = 453 \text{ nm} \) and, as can be seen in Figure 5.5, is very close to the peak position of the halo providing strong evidence that the halo is a result of constructive interference in the cavity formed by the recorded bit and the aperture.

The role of SPs was further confirmed by the absence of a halo in simulations where gold was replaced by a perfect electric conductor (PEC) to eliminate SP effects. SPs are not allowed to propagate on PEC/dielectric interfaces without surface structure because the electric field is excluded from the PEC interior [38]. Figure 5.5 shows the resulting simulated C-NSOM scans for an on resonance, \( \alpha = 50 \text{ nm} \) aperture in gold, an off resonance, \( \alpha = 50 \text{ nm} \) aperture in PEC and a retuned to resonance, \( \alpha = 75 \text{ nm} \) aperture in PEC. The halo represents \(~0.5\%\) increase in transmission for the gold case but is entirely missing from the PEC cases. Also plotted is the SP half-wavelength position showing the close match between the halo peak position and \( \lambda_{sp}/2 \). Owing to the fact that a recorded bit is required to produce the halo and the intensity of the SP wave is much smaller than the peak field produced by the C, the effect occurs only on readout and may be useful in increasing optical contrast and thus readout signal-to-noise ratios for encoding schemes that rely on edge transitions to store data (e.g., run length limited encoding).

![Figure 5.5 - FDTD simulated C-NSOM scans using real metal and PEC. The black line shows a simulated C-NSOM scan along the x-axis for an \( \alpha = 50 \text{ nm} \) aperture and a 100x100](image-url)

83
nm crystalline bit showing enhanced transmission at $x = \pm 400$ nm due to surface plasmon scattering. The red line shows the same aperture with the gold replaced by PEC. The decrease in the magnitude of the transmission dip is caused by the $\alpha = 50$ nm PEC aperture operating off its spectral resonance peak and is an expected result of the 0 nm skin depth of the PEC. The blue line shows a resized PEC aperture with $\alpha = 75$ nm that is resonant at 980 nm.

### 5.4.4 AFM Readout of Recorded Bits

To determine the physical size of the crystallized area, we removed the silicon nitride capping layer by etching the sample in 1:50 HF in water solution for 120 s and examined the Ge$_2$Sb$_2$Te$_5$ surface topography with an atomic force microscope (AFM; Digital Instruments Nanoscope 3000). Crystalline Ge$_2$Sb$_2$Te$_5$ is ~6.5% thinner than amorphous Ge$_2$Sb$_2$Te$_5$, causing surface depressions at the site of crystallization [109]. For a 20 nm film, we expect a change in height of ~1.2 nm for a fully crystallized film. Figure 5.3d shows an AFM scan of four bits recorded near threshold; Figure 5.3e and f show a cross section with FWHM of 53.5x50.2 nm indicating a potential data density of 223 Gbit/in$^2$. The transition between amorphous and crystalline is expected to be a step edge with minimal edge rounding and to accurately measure the size of the recorded bits, the AFM cross sectional data was fit to a two sided logistic of the form

$$y = \begin{cases} 
\frac{A}{1 + e^{-(x-x_c)/w}}; & x > x_c \\
\frac{A}{1 + e^{(x-x_c)/w}}; & x \leq x_c 
\end{cases}$$

(5.3)

where $A$ is the bit depth, $x_c$ is the bit center position and $w$ is the transition width of the bit edges. The fit to the bit identified by the cross hairs in Figure 5.6 is shown as a dashed red line in Figure 5.6b & c.
Figure 5.6 – (a) An AFM scan of bits recorded with 10 µs, 19.34 mW pulses. White arrows and dashed cross hairs indicate the position of recorded bits. (b,c) Cross sections along the bit identified in (d) by the cross hairs. Dashed lines show a two-sided logistic fit to the bit depression.

The behavior the optical readout response and bit size measured by AFM as a function of optical pulse power is plotted in Figure 5.7. Crystallized bit shape measured by AFM is asymmetric for larger doses but becomes nearly symmetric near threshold, as predicted by simulation. In contrast, the optical readout response FWHM values approach a width difference limit of ~70 nm. The width of the optical readout is much larger along the x-axis demonstrating that not only do the larger wings of the near-field spot in this dimension affect a change in the Ge₃Sb₂Te₅ during recording but they also interact with the resulting crystalline bit beyond the overlap of the bit with the nominal FWHM of the near-field optical spot. The result is a stacking of the effects and an optical response FWHM that is 50% larger in x than in y. This finding is also in agreement with the results of Chapter 4, where the effective width of the near-field probe was noticeably larger than the simulated FWHM. Simulated C-NSOM optical FWHM for crystalline bit sizes matching experimental AFM values are shown as dashed lines in Figure 5.7. These simulation results agree well with experiment; the ~10% difference in x-axis FWHM is likely due to ridge gap differences, caused by FIB spot size variation.
Figure 5.7 – Bit FWHM versus Optical Power. (a) The average bit size as measured by AFM and the average measured optical response FWHM for bits recorded with $t_{\text{pulse}} = 10$ µs and varying optical power. Dashed lines are FDTD simulated C-NSOM optical response FWHM for bits with x-y dimensions of 60×50, 80×60 and 90×70 nm (chosen to match the measured AFM values). Error bars are 1σ. Images (b) and (c) show example C-NSOM images of bits recorded at low and high power, indicated by the dotted ovals in (a). Images (d) and (e) show examples of the corresponding AFM images. Note the change of scale.

5.5 Summary

We have shown that C-apertures manufactured with our improved fabrication techniques and incorporated into a C-NSOM apparatus can overcome the low power transmission problems of conventional, aperture-based near-field data storage to record and read deeply subwavelength bits in Ge$_2$Sb$_2$Te$_5$ with bit sizes as small as 53.5×50.2 nm ($\lambda/20$) and offering a potential data density of 223 Gbit/in$^2$. During optical readout, a halo is observed around each bit that is shown via simulation to involve the propagation of surface plasmons along the exit surface of the metal layer. By calculating the SP wavelength on the exit surface we show that the halo peak occurs at a distance from the center of the aperture of roughly
indicating that the increased transmission is caused by constructive interference of the SP wave in the half wave cavity formed between the crystallized bit and the aperture.

Examining the bit shape by AFM, we show a progression from symmetric to asymmetric bits with increasing dose, in accordance with simulation. Further, we simulated the far-field optical response of the system showing superb agreement with experiment, proving the value of these simulations for designing future systems.

Finally, there are many prospective changes that can be made to greatly reduce the size of the recorded bit. First, alternative aperture shapes such as the elongation suggested by either Shi [49] or Sendur [50], or apertures with a sharpened ridge can further reduce the spot size and optical power requirements for data storage. There is a great deal of optimization available in the shape of the aperture and spot sizes of <10 nm are possible. Second, reducing the aperture size and operating at shorter optical wavelengths will reduce the bit size but may come at the cost of a slight reduction in efficiency due to the lower optical quality of most metals at shorter wavelengths. Third, reducing the layer thickness of the protective capping layer and air gap reduces the expansion of the near-field spot as it propagates away from the aperture exit. By applying these optimizations to aperture shape and media structure the C-aperture can provide an excellent approach to heat assisted magnetic recording (HAMR) and high density optical data storage in the future.
6 Conclusions

In this dissertation I have examined the real world capabilities of the C-aperture for use in near-field imaging and optical data storage applications. The C-aperture provides a superior method of producing a deeply subwavelength optical spot compared to the alternatives (e.g., solid immersion, sharp tips and aperture arrays). Specifically, it offers a smaller spot than can be achieved with SIL techniques by utilizing the lightning rod effect in a manner similar to that of sharp tips but without their large background signal. The folded interior shape of the aperture accommodates propagating and evanescent modes near cutoff and yields very high transmitted intensities by collecting light from an area of $\sim (\lambda/2)^2$. All of these features designate the C-aperture (and related apertures such as bowties and H-apertures) as the ideal choice for advancing subwavelength imaging and optical data storage.

Fabrication and utilization of the full capabilities of the C-aperture is difficult despite the promise of high performance predicted by simulations of C-apertures that use idealized pure noble metal optical parameters and idealized corners and sidewalls together. The primary barriers to high performance are structural non-idealities caused by DMM fabrication with a FIB and in Chapter 2 I introduced a method of fabricating apertures using silicon nitride membranes that masks the metal from surface rounding and gallium contamination. Realistic simulations of the TMM and DMM apertures show that DMM apertures perform significantly worse than the idealized case and that TMM fabrication largely restores the
aperture’s performance with a 630x increase in near-field intensity and a 2.2x decrease in optical spot size compared to the DMM case. Experiments comparing the far-field transmission properties of these TMM apertures to DMM apertures confirmed the simulation results, showing a significant increase in far-field transmission of >8.8x.

In Chapter 3, the ability of TMM C-apertures to produce a subwavelength spot capable of affecting a change in another material was demonstrated by utilizing two-photon absorption in a negative photoresist to acted as a high resolution probe of the C-aperture’s near-field. I showed that the near-field optical spot was confined to an area much smaller than the wavelength of light ($\lambda/8-\lambda/12$) and has an intensity at least 20x higher than the illumination light. I also showed that DMM and round apertures were not capable of polymerizing the photoresist at the low power levels used by the TMM aperture. This validation, if used in conjunction with the C-NSOM, proves the feasibility of high resolution photolithography using the C-aperture. If used as part of a multi-probe C-NSOM system, the approach could provide an alternative to current DUV/EUV lens-based lithography.

In Chapter 4, I present the first ever use of the C-aperture as a near-field imaging probe in the form of the C-NSOM instrument. Chrome nano-wires as small as $\lambda/20$ were clearly imaged with a resolution of $<\lambda/10$ and TE wires were found to exhibit a strong dipole coupling between the aperture and the chrome nano-wire that increased the transmitted intensity under certain conditions. This dipole coupling is not observed in TM wires. Also measured were chrome step edges that allowed the measurement of the resolving power versus tip-to-sample distance which was found to decrease by about 2 nm per 1 nm of $z$-propagation, in good agreement with simulation. The successful demonstration of the C-NSOM for imaging shows that this system can be used both as a general purpose high resolution scientific instrument and for defect inspection in semiconductor manufacturing.
Finally, in Chapter 5, I demonstrated the ability of the C-aperture to be used for optical data storage. The C-NSOM was shown to be capable of recording 53.5x50.2 nm ($\lambda/20$) data bits in a layer of Ge$_2$Sb$_2$Te$_5$ using a relatively low power of ~20 mW for a potential data density of 223 Gbit/in$^2$. The presence and size of the data bits was read both optically (using the C-NSOM) and with AFM scans of the Ge$_2$Sb$_2$Te$_5$ surface after recording. Additionally, I found evidence of surface plasmons propagating on the aperture film exit surface that result in an increased transmission halo surrounding each data bit upon optical readout and occurring at a distance from the center of the aperture of $\lambda_{sp}/2 = 453$ nm.

All of these simulations and experiments taken together show that the TMM C-aperture is capable of producing a deeply subwavelength spot with an effective size of $\lambda/10$ or better and a peak intensity that is at least 26x the intensity of the illumination light and 630x higher than the DMM aperture. Realistic simulation of the aperture shape and surrounding media has been used throughout this body of work and found to be in excellent agreement with the experimental results. The close correlation between simulation and experiment, both in the shape and intensity of predicted features should give us great confidence that FDTD tools can be used very effectively to design apertures for future systems with improved performance.

6.1 Future Work

By using FDTD tools for realistic modeling in conjunction with an understanding of the milling properties of the focused ion beam, it is possible to design and fabricate apertures that retain the high transmission of the aperture design used in this work, but exhibit a substantially reduced near-field optical spot size. Additional improvements in FIB technology (e.g., helium ion beam milling) will further the goal of fabricating apertures with small features and will allow the optimization of apertures for shorter wavelengths. Spot size
estimates of 5-10 nm are reasonable and would bring the resolution of the C-aperture to parity with the apertureless NSOM while retaining the small background signal inherent to apertures. Beyond optimizing the aperture shape, it is also possible to add structures to the aperture to improve the performance. A well known example of this is the addition of periodic corrugations around nano-apertures which have been shown to increase the near-field intensity by about a factor of ten. Another example, currently being studied in our group is the addition of a sharp metallic tip to the aperture ridge. This tip becomes, in effect, an apertureless NSOM tip that is sourced by the near-field spot of the C-aperture, further concentrating the spot through the lightning-rod effect.

In terms of practical application, a great deal of work remains. The behavior of the C-aperture when used for near-field inspection needs to be studied more generally, looking at alternative structures (e.g., closely spaced periodic arrays) and different materials (e.g., dielectrics). Perhaps most importantly, the scanning must be made massively parallel to make the imaging of large areas practical.

The application of C-apertures to optical data storage also offers many future opportunities. In addition to the optimization of aperture shape and layer thicknesses in the recording medium that I have already discussed, the integration of a C-aperture into a flying head and a demonstration of high speed recording would be significant steps. Also of interest would be studies on reflection mode read-out that would obviate the need for a collection apparatus on the side of the disk opposite the recording head.

Although there is still work to be done in maturing the use of C-apertures, I have taken the first steps towards their practical application. I have shown throughout this dissertation that the C-aperture is a good solution to many problems requiring a small, intense optical spot; by taking advantage of the available optimizations, the C-aperture can become a great solution.
7 Appendix

The purpose of these appendices is to cover the more mundane details of the experiments described in this dissertation for individuals that might be interested in duplicating and expanding on the work I have done. The two main areas where a detailed description is required are in simulation and experiment, particularly the specifics of the C-NSOM construction.

7.1 Finite Difference Time Domain (FDTD) Simulations

7.1.1 Background and Techniques

Finite difference time domain simulation of electromagnetic fields is a technique that has proved very useful in recent years due to the increasing computational power available. The method is well described elsewhere, particularly in Taflove and Hagness’s book “Computational Electrodynamics: the finite-difference time-domain method” so I will only give the specifics of the methods used for my work. I would like to point out that all of the C++ programming to implement the simulation (short of specifying the simulation space) was done by Paul Hansen and I am grateful to him for his fine work in this regard.

The simulation grid used for all the work in this dissertation was composed of a standard Yee cell (having no adaptive or graded meshing) although the grid was often meshed more finely in the z direction than in x and y because of the thin, z-stacked layers of material that needed to be resolved. There were 2 types of update equations used (excluding the
perfectly matched layer updates). First is the update equation for dielectrics that defines future electric field values in terms of their previous value and the curl of the magnetic field of adjacent points:

\[ E_{n+1/2}^n = c_1 E_{n-1/2}^n + c_2 \cdot (\nabla \times H^n - J_{\text{source}}^n) \]  

(7.1)

where \( n \) is the index of the time step, \( \Delta x, \Delta y, \) and \( \Delta z \) are the grid sizes from the Yee cell and define the distance over which the finite difference curl of \( H \) is taken and \( J_{\text{source}} \) is any current source term forcibly added to the simulation. The constants are defined as:

\[ c_1 = \frac{1 - \sigma \Delta t}{2\varepsilon} \]  

(7.2)

and

\[ c_2 = \frac{\Delta t}{\varepsilon} \]  

(7.3)

where \( \sigma \) is the material’s conductivity at the \( E \)-field point being calculated, \( \varepsilon \) is the permittivity also taken at the \( E \)-field point being calculated and \( \Delta t \) is the time step specified in the simulation. The update equation for the magnetic field is:

\[ H_{n+1}^n = b_1 H_{n}^n - b_2 \cdot (\nabla \times E_{n+1/2}^n - M_{\text{source}}^{n+1/2}) \]  

(7.4)

where \( M_{\text{source}} \) is a magnetic field source term and the constants are:

\[ b_1 = 1 - \frac{\sigma^* \Delta t}{2\mu} \]  

(7.5)

and

\[ b_2 = \frac{\Delta t}{\mu} \]  

(7.6)
where $\sigma^*$ is a magnetic loss term not used in the simulation and $\mu$ is the material’s permeability.

The material model used for the metals in all of these simulations is the Drude model. The dispersive permittivity of a Drude metal is described in the simulations using the equation

$$\varepsilon(\omega) = \varepsilon_\infty - \frac{\omega_p^2}{\omega^2 + i\Gamma\omega} \quad (7.7)$$

where $\varepsilon_\infty$ is the permittivity at infinite frequency, $\omega_p$ is the plasma frequency, $\Gamma = 1/\tau_c$ and $\tau_c$ is the relaxation time constant. The update equations are:

$$E|^{n+1/2} = E|^{n-1/2} + \frac{\Delta t}{\varepsilon_\infty} (\nabla \times H|^{n} - J|^{n}) \quad (7.8)$$

$$H|^{n+1} = H|^{n} + \frac{\Delta t}{\mu} (\nabla \times E|^{n+1/2}) \quad (7.9)$$

where the current term is now a reactive term defined by:

$$J|^{n+1} = J|^{n} \left(\frac{2}{2 + \Gamma\Delta t}\right) + E|^{n+1/2} \left(\frac{2\varepsilon_0\omega_p^2\Delta t}{2 + \Gamma\Delta t}\right) \quad (7.10)$$

The plane wave excitation was introduced using the Total-field/Scattered-field (TFSF) formulation outlined in §5.8.1 and §5.8.2 of Taflove & Hagness [86].

### 7.1.2 The Simulation Space

The simulation space and any batch simulations were generated using a MatLab script to build the necessary structures and issue system commands to run the C++ executable. The simulation space consisted of a 3D box, the interior of which constituted the simulation space. The simulation space was surrounded on all sides by a perfectly matched layer (PML) 10 cells thick, the purpose of which was to absorb any electromagnetic waves leaving the simulation space and thus prevent unwanted reflections. The box defining the simulation space was 1.15 $\mu$m on a side in x and y which, for a 5 nm grid, makes the simulation size 230x230 cells in x and y. In the z direction, the number of cells was determined by the number and thickness of...
the simulated layers. Typically, the layers included a silicon nitride membrane layer ~100 nm thick, a gold layer ~100 nm thick, an air gap ~10 nm thick, and any media or sample layers. The buffer between the simulated materials and the edge of the simulation in this direction was set to 25 cells to properly absorb any near-fields.

The TFSF boundary was placed 5 cells interior to the PML boundary and the electric field intensity of the incident plane wave excitation was defined by:

$$E(t) = \sin(\omega_{src}(t - \delta)) \cdot e^{\frac{(t-\delta)^2}{w^2}}$$

(7.11)

where $\omega_{src}$ is the base frequency of the source (the free space frequency of a $\lambda = 980$ nm wave), $t$ is time, $\delta = 3w$ is a pulse delay term, and $w$ is the width of pulse envelope (set to $0.5 \times 10^{-15}$ s).

Figure 7.1 shows an example cross section through the simulation space. The plane wave is sourced from the top TFSF boundary.

Figure 7.1 – Example cross section of the simulation space taken along the x axis and through the aperture ridge. This simulation happens to be from Chapter 5 and is of a 60x50 nm crystalline bit slightly offset in x from the aperture center.

As previously mentioned, the aperture shape is created by convolving the idealized, sharp cornered mill pattern (found in Figure 2.2a and others) with a gaussian with a 33 nm FWHM to produce a depth map. The resulting surface is scaled so that the maximum is equal to the thickness of the silicon nitride membrane plus gold layers and then that depth map is
multiplied by the overmill factor used in the experiment (typically 1.5-2x). The result is shown in Figure 7.2 where any points in the depth map that are deeper than the bottom surface of the gold layer are represented as pure white, and the top silicon nitride surface is shown as black. Other depths are represented as grayscale.

![Figure 7.2](image)

Figure 7.2 – The height map for an $\alpha = 50$ nm aperture used for “milling” the aperture in the simulation.

7.1.3 Analysis

A. Simulation Output

The raw output from the simulation consists of electric and magnetic field amplitudes for each cell in the simulation and at each time step for the whole simulation. Because most of the simulations were run for 15000 time steps, it is impractical to store all of the information produced by the simulation and most of the electromagnetic field data is dumped after it is no longer needed by the simulation to calculate future time steps. Data is saved at certain planes through the simulation that are of particular interest. Common planes of interest are x-z and y-z planes through the center of the aperture and x-y planes after the aperture exit spaced at various distances away from the exit surface. The x-y planes are most often recorded one or two cells from the exit surface and at some other plane of interest (e.g., in the optical data storage case, the other planes are inside the Ge$_2$Sb$_2$Te$_5$ layer). Unfortunately, even when
restricting the saved data to a handful of individual planes, it is still helpful to sparse the time
data by 15 steps (i.e., data is stored only every 15th time step). Figure 7.3 shows an example
of the electric field amplitude at a single point in the simulation versus time. This data is
stored for each Yee cell in the recorded planes and serves as the exclusive source of
information for all of the analysis that follows.

![Electric Field Amplitude](image.png)

**Figure 7.3 – Example plot of the electric field amplitude at the aperture exit versus time.**

**B. Electric Field Distributions**

To calculate the intensity of the electric field at a particular wavelength or free-space
frequency, the field amplitude versus time data must be decomposed into its various frequency
components. The data collected during the simulations is actually sparse in information at the
frequencies we are interested in; they contain much more high frequency information than we
need and would need to be run for a much longer time to produce high resolution data at the
comparatively long wavelengths we are using. As a result, simple FFT decomposition into
frequency components is unacceptably sparse for wavelengths around 1 µm and I have instead
used a DFT to calculate the frequency component at each wavelength of interest. It is
important to realize that this doesn’t actually create any new information, but instead amounts
to an interpolation that uses the FFT of the time window as the interpolation spline.
Using the DFT, I first build a transmitted intensity curve by calculating the transmitted fields for a wide range of wavelengths at a single point in the simulation (centered in x-y and immediately after the aperture in z). I also calculate the electric field amplitude and phase for each of the three electric field components at each point in the recorded planes for a particular wavelength of interest (typically 780 or 980 nm). Additionally, the amplitude of the source wave is recorded (at a single point, since it is a plane wave and uniform throughout the wavefront) and the electric field amplitude of the source wave at various wavelengths of interest is calculated to provide a normalization value. From these values, the electric field intensity can be calculated straightforwardly as

\[
I = \frac{\left( |E_x|^2 + |E_y|^2 + |E_z|^2 \right)}{|E_{src}|^2}
\]

(7.12)

This value is used for almost all the electric field intensity plots in this dissertation and is often referred to as the near-field enhancement because it is normalized to the source wave intensity.

C. Far-Field Transmission

Calculation of the far-field transmitted power is a more complicated endeavor. Because we cannot run a simulation large enough that we can measure the far-field directly (in the way we measure the near-field amplitudes), it is necessary to devise a method of calculating the far-field from the near-field amplitudes. Fortunately, this is a solved problem because of its obvious applicability to antenna theory. Taflove & Hagness cover this in detail in Chapter 8 of their text. I have used their described method utilizing the surface equivalence theorem to calculate the far-field transmitted intensity as a function of emission angle. The fields used in the calculation are the \( E \) and \( H \) fields measured at a single x-y plane occurring at a z position after any interacting structures. The other bounding surfaces (x-z and y-z planes) are ignored under the assumption that their contribution is negligible to the +z emission if the x-y plane is taken close enough to the emitting surface. For the plain gold on silicon nitride
structure the measurement plane is anywhere after the gold surface, for the membrane tip examining chrome nano-wires the measurement plane is anywhere in the glass substrate, and for the membrane tip over Ge2Sb2Te5, the measurement plane is anywhere after all the layers of the recording stack and inside the glass substrate. The equations used to calculate the radiation cross section (RCS) start with the equivalent virtual surface electric and magnetic currents:

\[ \mathbf{J}_s = \hat{n} \times \mathbf{H} \]  
\[ \mathbf{M}_s = -\hat{n} \times \mathbf{E} \]

Where \( \mathbf{H} \) and \( \mathbf{E} \) are the phasor field quantities measured at the exit plane and \( \hat{n} \) is the surface normal, \( \hat{z} \) in our case. From these surface currents, the following interim quantities are calculated for each angle \((\theta, \phi)\) of interest:

\[ \bar{N}_\theta = \iint_S (J_x \cos \theta \cos \phi + J_y \cos \theta \sin \phi) e^{ikr'\cos \psi} ds' \]  
\[ \bar{N}_\phi = \iint_S (-J_x \sin \phi + J_y \cos \phi) e^{ikr'\cos \psi} ds' \]  
\[ \bar{L}_\theta = \iint_S (M_x \cos \theta \cos \phi + M_y \cos \theta \sin \phi) e^{ikr'\cos \psi} ds' \]  
\[ \bar{L}_\phi = \iint_S (-M_x \sin \phi + M_y \cos \phi) e^{ikr'\cos \psi} ds' \]

where \( k \) is the wavevector of the wavelength of interest and

\[ r' \cos \psi = x' \sin \theta \cos \phi + y' \sin \theta \sin \phi \]

where \( x' \) and \( y' \) are the spatial positions in the x-y plane of the measured field components and \( ds' = dx'dy' \). These integrals become sums over the available x-y measurement plane for each \((\theta, \phi)\) point of interest, which practically runs over the range \( \theta = (0, \pi) \) and \( \phi = (0, 2\pi) \), thus spanning the exit half-space. The resulting radiation pattern becomes
\[ \text{RCS}(\theta, \phi) = \frac{k^2}{8\pi \eta_0 P_{inc}} \left( |L_\phi + \eta_0 \bar{N}_\phi|^2 + |L_\theta - \eta_0 \bar{N}_\theta|^2 \right) \] (7.20)

where \( \eta_0 = \sqrt{\mu_0 / \varepsilon_0} \) and

\[ P_{inc} = \int \int_{\mathcal{S}} \frac{\varepsilon_0 C}{2} |E_{src}|^2 \, ds' \] (7.21)

An example RCS pattern is shown in Figure 7.4 and exhibits strong emission lobes in the \( \phi = 0 \) and \( \pi \).

Figure 7.4 – An example RCS pattern for an \( \alpha = 50 \) nm aperture over a-Ge\(_2\)Sb\(_2\)Te\(_5\).

Finally, to calculate the signal collected by our apparatus, the RCS is integrated over the full range of \( \phi \) and over the range of \( \theta \) covered by the numerical aperture of the collection optics (0.25 for most of this work). The intensity in the far field is then

\[ I_{FF} = \int_0^{\theta_{NA}} \left( \int_0^{2\pi} \frac{\text{RCS}(\theta, \phi)}{2\pi} \sin \theta \, d\phi \right) \, d\theta \] (7.22)

where \( \theta_{NA} \) is the maximum angle included in the numerical aperture of the collection optics. This number, \( I_{FF} \) (normalized to a baseline value), is used to build all the v-NSOM plots in this dissertation. Each value in the v-NSOM plots represents a wholly different simulation.
with the sample translated in the simulation as it is in the experiment. In addition to
normalization to baseline, all of the $I_{FF}$ values are first modified by adding the contribution of
light transmitted through the gold layer from areas not included in the near-field calculation.
This additional contribution is calculated by using a thin film transmission calculation and
integrating the power transmitted, multiplied by a 2D gaussian profile of mode field diameter
2.1 $\mu$m (the spot produced by the illuminating fiber). The 1.15 $\mu$m square included in the
near-field calculation is subtracted from the integrated sum to arrive at the contribution from
this effect.

7.2 C-NSOM Operation

7.2.1 Mechanical Construction

The C-NSOM is configured as an inverted microscope with a fiber NSOM positioned
on top (see Figure 7.5). The distance between fiber tip and sample is shear force feedback
regulated using an electrically excited quartz tuning fork (Digikey part number X-801ND)
operating near its design frequency of 32.768kHz. In order to mount the tuning fork and
optical fiber, the tuning fork is first extracted from the metallic cylinder by carefully crushing
the epoxy seal at the cylinder’s base with pliers and removing the quartz tuning fork. The bare
tuning fork is then glued to a plastic mount with NOA-81 and a cleaved optical fiber is glued
to one prong of the tuning fork (also with NOA-81). The tuning fork is electrically excited on
one prong with a 6 mV peak-to-peak sine wave and the current excited in the opposite prong is
input into a two stage, 1x10^6 gain transimpedance amplifier located as close as possible to the
tuning fork. The resulting voltage signal is then fed to a lockin amplifier to obtain the
magnitude of the excited oscillation. The tip-sample distance is controlled by a piezo actuator
stack (Thorlabs P/N AE0203D08F) with a 0-100 V throw of 6.1 $\mu$m that has been
incorporated into a 1D box flexure having a resonant frequency modeled at 1.3 kHz. The structure is shown in Figure 7.6.

Figure 7.5 – The C-NSOM setup showing the components external to the tip sample interface. Note that the setup shown here is rotated 90° counter-clockwise from the inverted microscope configuration of the experiment.

Figure 7.6 – Z-axis flexure stage. The piezo stack is inserted in the u-shaped gap forcing the open box to deform and translate down. The tuning fork is mounted at the end of this structure.
7.2.2 Tip Assembly

Connecting the membrane with truncated pyramidal tip to the end of an optical fiber is difficult. The assembly begins after the optical fiber has been glued to the quartz tuning fork and starts with placing the membrane (supported by the silicon frame) on the sample stage below the fiber with the pyramidal tip facing down. The cleaved fiber facet is first made parallel to the membrane surface by using a Linnik phase shifting interferometer that images the fiber end to measure the average tilt of the cleaved fiber facet. The tilt of the membrane surface is also measured and the sample stage tip-tilt is adjusted so that the fiber facet and the membrane surface are parallel. Figure 7.7 shows the next steps. The cleaved fiber facet is never perfectly flat and is generally saddle shaped with a peak-to-peak height difference of ~1 µm. Because of this, if the core of the fiber and the truncated pyramid are aligned in x-y and the glue is cured, the large glue thickness variation leads to non-uniform glue shrinkage (UV cure adhesives typically shrink by ~1-2%) which pulls the fiber laterally relative to the membrane and misaligns the core and pyramid. To overcome this difficulty, a small quantity of glue is applied to the fiber tip and the fiber tip is brought into contact with the membrane in an area backed by the silicon frame (Figure 7.7a-b). Capillary forces move the glue into any gaps and after curing, the fiber tip and membrane are separated by pulling on the fiber, leaving a planarized, perfectly parallel surface on the end of the fiber. A very small amount of glue is then applied to the planarized surface and the fiber core is aligned in x-y to the truncated pyramid by examining the symmetry of the transmitted light (Figure 7.7d). Finally, the glue is cured and the assembled tip is freed from the silicon frame by first forcing the fiber tip down to break the membrane loose at the perimeter and then retracting the assembled tip from the frame (Figure 7.7e).
Figure 7.7 – (a) A droplet of glue is placed on the end of the cleaved fiber facet. (b) The fiber tip is touched to the membrane frame surface away from a membrane and the glue moves to fill any gaps. (c) After UV curing the glue, the tip and glue are snapped free of the surface. (d) A very small amount of additional glue is added to the tip, the fiber core and pyramidal tip are aligned and the glue is UV cured. (e) The membrane is broken free from the silicon frame.

7.2.3 Scanning Methodology

A diagram of the amplitude of the excited oscillation as a function of frequency is shown in Figure 7.8a. The amplitude peaks at the resonant frequency of the tuning fork and fiber structure with a value of $V_{\text{max}}$. The signal reaches an amplitude, $V_{\text{min}}$, when far detuned from the resonant frequency that is non-zero because of the capacitive leakage through the tuning fork. For operation of the feedback loop, the frequency of the 6 mV sine wave is detuned from the maximum amplitude point to a lower frequency, $f_0$, such that $V_0$ is 90% of $V_{\text{max}}$. The purpose of this detuning is because at the peak, the ring-up and ring-down time constants for the tuning fork are substantially different making it very difficult to tune the PI control loop for maximum stability. At the 90% of $V_{\text{max}}$ point, the two time constants are closer to each other and the PI loop is easier to tune.
Figure 7.8 – (a) Tuning fork oscillation amplitude versus frequency. (b) Tuning fork oscillation amplitude versus tip-sample distance.

Figure 7.8b plots the oscillation amplitude versus tip-sample distance. The out of contact amplitude starts at the $V_0$ value and is first damped by tip-surface interactions beginning at a separation of about 20 nm [101]. The set point for the feedback loop is set to 80% of $V_0$.

The feedback loop is a simple proportional-integral (PI) control loop and is implemented using MatLab’s Realtime PC software. Differential feedback is not used. This digital feedback loop is time discrete and updates the output to the z-distance high voltage controller every 200 $\mu$s. Gain settings are typically $P = 0.02$ and $I = 0.7$ although variable tip-sample interaction strength may require that these values be adjusted. When the feedback loop is engaged, the output to the high voltage amplifier is increased, extending the z-piezo until the tip-sample interaction damps the tuning fork oscillation and the oscillation amplitude matches the $V_{set}$ value. The feedback loop then maintains the tip-sample distance such that the oscillation amplitude is constant. With the feedback loop engaged and the tip near the surface, AFM data can be collected by recording the magnitude of the voltage sent to the z-piezo. As the x-y position is changed, the feedback loop adjusts this voltage to maintain a constant distance between tip and sample and the AFM height is defined as

\[ h = -10V_{feedback} \cdot \left( \frac{6.1 \ \mu m}{100 \ V} \right) \]  

(7.23)
where the feedback voltage has been multiplied by ten (because of the high voltage amplifier gain) and the conversion ratio of the piezo stack. All the C-NSOM data in this dissertation was collected using this method and the straightforward collection of transmitted light for the optical measurements.
8 Bibliography

42. J. Helszajn, Ridge waveguides and passive microwave components (The Institution of Electrical Engineers, 2000).
67. M. J. Vasile, J. Xie, and R. Nassar, "Depth control of focused ion-beam milling from a numerical model of the sputter process," in Papers from the 43rd international conference on electron, ion, and photon beam technology and nanofabrication, (AVS, 1999), 3085-3090.


